Atomistic simulations of cavitands, cytochrome P450, and natural product binding to actin

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Tim Knehans
aus
Brandenburg a.H.
DE

Promotionskomitee
Prof. Dr. Amedeo Caflisch (Vorsitz)
Prof. Dr. Cristina Nevado

Zürich, 2015

Summary

Guidance of experiments and investigation of experimental outcomes through computational models is increasing in both scope and importance. Few modern research topics can be tackled with laboratory experiments alone as these rarely achieve the desired resolution or would be combinatorically too vast to apply exhaustively. This thesis presents four cases in which computational models either provided an interpretation of experimental outcomes or guided experiments in order to efficiently achieve the intended outcome. The significant difference in function of the target molecules presented herein underlines the strength of computer-based models to aid and guide a large variety of experiments.

The first case is the molecular dynamics analysis of medium-sized organic molecules known as cavitands, which display two distinct conformations (an opened kite and a closed vase conformation) as a function of the physiochemical environment. High precision inter-dye distance distributions as well as relative dye-dye angles (κ^2) were obtained from molecular dynamics simulations. By incorporating these distances and κ^2 values into a simple Markov state model, a Förster Resonance Energy Transfer (FRET) experiment was emulated, giving modeled fluorescence anisotropy decay and donor fluorescence decay curves that highly correlated with experimentally determined FRET signals.

The cytochrome P450 3A4 variant (CYP3A4) is the most abundant cytochrome P450 in the human liver. Its high promiscuity leads to CYP3A4 processing nearly every xenobiotic entering the human body. We set out to model the binding behavior of the anticonvulsant carbamazepine (CBZ) for which no X-ray crystal structure or NMR ensemble in complex with CYP3A4 is available. Utilizing molecular dynamics simulations in conjunction with cluster analysis implemented in the program WORDOM, we proposed target residues for mutational studies. Compared to the wild type, one mutant (A370V) displayed a significantly higher turnover rate at high substrate concentrations while another mutant (I369F) saw increased turnover at low substrate concentrations. Further, the kinetics of CBZ epoxidation of the A370V mutant differed from the standard Michaelis-Menten kinetics of wild-type CYP3A4. Its sigmoidal kinetic profile indicated induced cooperativity, which was likewise observed for two other proposed mutants (S119A, A370L). In addition, all single-point mutants (A370L, A370V, I369F, I369L, and S119A) showed increased spin-shift fractions in comparison to wild-type CYP3A4, indicating a shift of the spin equilibrium towards the ferric high-spin state. Finally, a

single simulation showed the egress of a CBZ molecule via a channel previously hypothesized in the literature.

The importance of epigenetics in the time-dependent development of cells and its implication in cancer has recently come to light. Resultant efforts to disrupt the action of epigenetic reader molecules called bromodomains, which provide an interface for large protein complexes to interact with chromatin, have led to a series of chemical probes and drugs. The SMARCA4 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4) bromodomain has been indicated as a potential target for certain SWI/SNF-mediated cancers. Applying an established virtual screening scheme, Anchor-based Library Tailoring Approach (ALTA), in combination with the novel docking software rDock and an interaction energy-based scoring scheme, we identified a trans-glycoluril-containing compound out of a multimillion molecule database (ZINC). This compound was verified *in vitro* by an ALPHA screen assay, resulting in an IC_{50} value of 20 µM.

Actin is the most abundant protein in higher organisms and is the primary subunit of actin filaments, which are crucial in maintaining stability and motility of the cell its components. A series of natural products extracted from deep sea sponges has been identified as altering the dynamics of actin polymerization. The two primary sites of interference with actin polymerization are the barbed end and, with the characterization of the natural product latrunculin, the inter-domain region between subdomains 2 and 4. Recently, iriomoteolide 3a was shown to impact the dynamics of actin polymerization, though its mechanism was unknown. A 7 µs molecular dynamics simulation of actin with iriomoteolide 3a is presented herein. These simulations were investigated for the interaction frequency of iriomoteolide 3a with each actin residues in order to identify preferred interaction hot spots. Following the identification of the barbed end as the most likely site of association, two putative binding modes of iriomoteolide 3a with monomeric actin are presented while simulations of the natural product reidispongiolide A support its mechanism of actin filament severing as proposed in the literature.

Zusammenfassung

Die Anleitung von Experimenten und die Auswertung experimenteller Resultate durch computergestützte Modelle breiten sich ständig aus, sowohl in ihrem Umfang, als auch in ihrer Wichtigkeit. Wenige moderne experimentelle Fragestellungen können heute noch ausschliesslich mit Laborexperimenten untersucht werden, da diesen entweder die nötige Auflösung fehlt, oder sie kombinatorisch zu anspruchsvoll sind, um sie umfangend einzusetzen. Diese Dissertation präsentiert vier Anwendungsfälle, in denen computerbasierte Modelle entweder die Interpretation der experimentellen Daten ermöglichten, oder Experimente anleiteten, um das vorher gesetzte Ziel effizient zu erreichen. Die signifikanten Unterschiede in den Funktionen, der hier präsentierten Zielmoleküle, betont die Stärke von computerbasierten Modellen, eine grosse Auswahl an Experimenten zu leiten und zu unterstützen.

Der erste Anwendungsfall ist die molekulardynamische (MD) Analyse von mittelgrossen organischen Molekülen, die Cavitanden genannt werden, und zwei verschiedene konformelle Zustände aufweisen (eine offene Drachen- und eine geschlossenen Vasenform), welche eine Funktion der physikochemischen Umgebung sind. Hochpräzise Distanzverteilungen, ebenso wie relative Farbstoff-Farbstoff-Winkel (κ^2), wurden mit Hilfe von MD Simulationen bereitgestellt. Durch den Einarbeitung dieser Distanzen und κ^2-Werte in ein einfaches Markov-Modell, wurde ein Förster Resonanz Energie Transfer (FRET) Experiment emuliert, welches modellierte Fluoreszenzanisotropiezerfall- und Donorfluoreszenzzerfall-Kurven generierte, welche stark mit denen, im Labor durchgeführten, FRET-Experimenten korrelierten.

Die Zytochrom P450 3A4 Variante (CYP3A4) ist das am häufigsten vorkommende Zytochrom P450 in der menschlichen Leber, dessen hohe Promiskuität zur Umsetzung von fast allen Fremdkörperstoffen führt, die in den menschlichen Körper gelangen. Wir begannen mit der Modellierung des Bindeverhaltens von Carbamazepine (CBZ), einem Antikonvulsivum, für welche weder Röntgen-Kristall-Strukturen, noch Nuclear Magnetic Resonance (NMR) Ensembles im Komplex mit CYP3A4 vorhanden sind. Durch die Verwendung von Moleküldynamiksimulationen in Kombination mit Clusteranalyse, implementiert im Programm WORDOM, schlugen wir Aminosäuren für nachfolgende Mutationsstudien vor. Im Vergleich zum Wild-Typ, wies eine Mutante (A370V) eine signifikant höhere Umsatzrate bei höheren Substratkonzentrationen auf, während eine andere Mutante (I369F) eine höhere Umsatzrate bei niedrigeren Substratkonzentrationen zeigte. Weiterhin wich die Enzymkinetik der CBZ-

Epoxidierung der A370V-Mutante merklich von der der Standard-Michaelis-Menten-Kinetik ab, welche das Wildtyp-CYP3A4 zeigt. Das sigmoidale Kinetikprofil wies auf eine induzierte Kooperativität hin, welche weiterhin für zwei andere Mutanten beobachtet wurde (S119A, A370L). Zusätzlich zeigten alle Einpunktmutanten (A370L, A370V, I369F, I369L, S119A) erhöhte spin-shift Fraktionen im Vergleich zu Wildtyp-C3A4, welche auf einen hoch-spin Eisenzustand hinweisen. Zuletzt zeigte eine Simulation das Verlassen eines CBZ-Moleküls, durch einen Kanal, welcher bereits früher in der Literatur postuliert wurde.

Die Gewichtigkeit epigenetischer Prozesse in der zeitabhängigen Entwicklung von Zellen und deren Implikationen in Krebs, rückten erst kürzlich in das Rampenlicht. Die darauf zurückzuführenden Bestrebungen epigenetische Lesemodule, sogenannte Bromodomänen, welche in grossen Proteinkomplexen die Interaktion mit Chromatin ermöglichen, zu stören, führten zu einer Serie chemischer Sonden und Medikamente. Die SMARCA4 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4)-Bromodomäne wurde als potentielles Ziel für bestimmte SWI/SNF Krebsarten identifiziert. Durch die Anwendung einer etablierten virtuellen Screeningmethode, Anchor-based Library Tailoring Approach (ALTA), in Kombination mit dem neuen Dockingprogramm rDock und einem interaktionsenergiebasiertem Bewertungsschema, haben wir einen trans-glycoluril beinhaltenden Stoff aus einer multi-Millionen grossen Datenbank (ZINC) identifiziert. Dieser Stoff wurde *in vitro* durch eine ALPHA Screen Prüfung auf 20 µM IC50 verifiziert.

Actin ist das, in höheren Lebewesen, am häufigsten auftretende Protein und ist die primäre Untereinheit von Acinfilamenten, welche unabdinglich sind, für die Instanthaltung und Bewegungsleistung der einzelnen Zellbestandteile. Eine Serie an Naturstoffen, welche von Tiefseeschwämmen extrahiert wurde, weist verändernde Eigenschaften von Actinpolymerisierungen auf. Die zwei primären Interaktionsstellen für Interferenz mit Actinpolymerisierung sind das "barbed end" und, seit der molekularbiologischen Charakterisierung von des Naturstoffes Latrunkulin, auch die Interdomänenregion zwischen den Unterdomänen 2 und 4. Erst kürzlich wurden Einflüsse von Iriomoteolide 3a auf Actinpolymerisierung nachgewiesen. Es wird eine7 µs MD-Simulationsserie von Iriomoteolide 3a mit monomerischem Actin präsentiert. Diese Simulationen wurden auf die Interaktionsfrequenz von Iriomoteolide 3a mit jeder Aminosäure überprüft. Nachdem das "barbed end" als der wahrscheinlichste Hauptinteraktionspunkt identifiziert wurde, erfolgte die Präsentation von zwei möglichen Bindeposen. Simulationen mit dem Naturstoff

Reidispongiolide A bestärken den vermuteten Interaktionsmechanismus mit dem Actinfilament.

Contents

1 Introduction

5.1 Introduction

5.2 Methods

5.3 Results

Chapter 1

Introduction

1.1 Computer-based simulation as an aid to experimentalists

The investigation of biomolecular processes is located at the very basis of research which leads to improved conditions of human life. With each piece of knowledge gained on biological pathways and the intricacies of cells, the chance to develop a cure for a related disease increases. Today, few experiments can be conducted without the support of computer-based modelling. If the experimental observables can be compared or coupled to the parameters of modeled systems, the former might be explained on the level of detail of the latter[1]. In the field of biomolecular research, molecular simulations have become an important tool to facilitate and extend the insights gained from experimental properties such as X-ray diffraction intensities, NMR parameters, quantities derived from EPR, CD infrared, RAMA or fluorescence spectroscopy[2–6]. With increases in computational resources, the scope of simulations grows proportionately to the complexity of simulations as well as the breadth of natural phenomena they aim to explain. Studies of polypeptide, folding into their native protein conformation, as well as aggregation of proteins into oligomers and lipids into micelles are now among the most commonly performed simulation studies[7,8].

With the development of force fields for small organic compounds, high-resolution investigations into protein substrate interactions can be conducted on relevant time scales. Simulations which model the unbinding of small molecular fragments have been shown to be in agreement with experiments and can be applied as a metric to identify potent from inactive small molecular inhibitors[9]. The most attractive property to be obtained from molecular simulation of biomolecules and small organic compounds is the estimation of biding free energies by molecular dynamics simulation[10]. Although the accuracy of this method is highly dependent on sampling time, simplified models have provided acceptable results[11,12].

1.2 Molecular dynamics simulations

Molecular dynamics simulations in particular have become a very important tool to facilitate the interpretation of experimental biological data on an atomistic level[1]. Its significance was most recently recognized by awarding the 2013 Nobel Prize in chemistry to Martin Karplus, Michael Levitt, and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*[13]. Molecular dynamics simulations represent atomic motions in terms of classical ball-on-a-spring mechanics in which the bonded energy is generally approximated by the sum of following terms:

$$E_{bond} = \sum_{\boldsymbol{n}} k_b (r - r_0) \tag{1}$$

where $\mathbf{E_{bond}}$ is the bond energy calculated over all covalent bonds **n** with a harmonic potential $\mathbf{k_b}$ and the difference between given bond length r and equilibrium bond length $\mathbf{r_0}$.

$$E_{angle} = \sum_n k_\theta (\theta - \theta_0)^2 \tag{2}$$

in which the angle energy $\mathbf{E_{angle}}$ is the sum over all products of the stiffness term $\mathbf{k_\theta}$ and the difference between current angle $\boldsymbol{\theta}$ and reference angle $\boldsymbol{\theta_0}$ squared.

$$E_{torsion} = \sum_m A[1 + \cos(n\tau - \phi)] \tag{3}$$

where torsional energy term $\mathbf{E_{torsion}}$ is expressed as a periodic function in which **ϕ**, **A**, and **n** are empirically determined values describing shift, amplitude, and periodicity of the cosine function for each every dihedral **τ** of four consecutively bonded atoms. Additional to these covalent parameters, non-covalent parameters play a crucial role in determining the correct interaction between atoms. As atoms in molecular dynamics simulations are approximated as point charges with a van der Waals radius, the non-bonded interactions are represented by:

$$E_{Coulomb} = \sum_i \sum_j \frac{qiqj}{4\pi\varepsilon_0 r_{ij}} \tag{4}$$

where the electrostatic contribution to the total energy of a system $\mathbf{E_{Coulomb}}$ is modelled as a Coulomb potential in which $\mathbf{q_i}$ and $\mathbf{q_j}$ are two point charges at distance $\mathbf{r_{ij}}$ while $\boldsymbol{\varepsilon_0}$ is the dielectric constant of vacuum.

$$E_{vdW} = \sum_i \sum_j (\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}}) \tag{5}$$

the Lennard-Jones potential used to describe the van der Waals (vdW) contributions to the energy of the system contains an attractive term $\mathbf{1/r^6}$ and a repulsive term $\mathbf{1/r^{12}}$ while $\mathbf{B_{ij}}$ and $\mathbf{A_{ij}}$ describe the location and depth of the minimum of the potential for two non-covalent atoms **i** and **j**.

$$V = E_{bond} + E_{angle} + E_{torsion} + E_{Coulomb} + E_{vdw} \tag{6}$$

Summing over all terms, $\mathbf{E_{bond}}$, $\mathbf{E_{angle}}$, $\mathbf{E_{torsion}}$, $\mathbf{E_{Coulomb}}$, $\mathbf{E_{vdw}}$ describes the potential energy **V** as a function of atom positions.

To propagate the system through time, Newton’s equation of motion is then solved for the system of **N** interacting atoms:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = ma_i = F_i, i = 1 \dots N \tag{7}$$

Where $\mathbf{m_i}$ and $\mathbf{a_i}$ are the mass and acceleration of atom **i** while the force $\mathbf{F_i}$ acting on atom **i** is the negative gradient of the potential energy:

$$F_i = -\frac{\partial V}{\partial r_i} \tag{8}$$

Over the years, a variety of force fields have been generated for proteins, nucleic acids, and sugars out of which CHARMM, AMBER, OPLS and GROMOS represent the most prominent ones[14–18]. Biomolecules like proteins and nucleic acids provide an intrinsic advantage for the parameterization effort as they are made up of a small amount of subunits like amino acids or nucleic bases which can be relatively easy parameterized to create an accurate description of the whole system as a combination of its subunits. However, more and more emphasis is put on the interaction of small organic molecules with their respective protein or DNA target. The combinatorial complexity of small organic compounds provides a greater challenge as exhaustive *ab initio* parametrization would take considerable amounts of time and resources. Thus, in the past some atomic force fields for small molecules were derived for a simplified view of molecules where for instance distinctions between carbon atoms were only made based on the number of bonds they provide with their immediate neighbors and if they are in an aromatic system or not, exemplified by the TRIPOS Atomic Force Field[19]. However, to provide more accurate parameters, a series of more complex force fields were developed. The Merck Molecular Force Field (MMFF94) was one of the most wildly applied force fields due to its ability to reproduce properties of small molecules in condensed phase[20]. Subsequently, efforts have been made to generate small molecular force fields for other protein force fields such as the General Amber Force Field (GAFF) and the OPLS-AA force field[21,22]. Most recently a general force field for CHARMM (CGenFF) was developed with the clear goals to provide CHARMM-consistent parameters and an ongoing parametrization effort in order to obtain the means to accurately predict the interaction of small molecular ligands and their protein targets[23]. CGenFF provides a series of explicitly parameterized fragments which represent the basis to parameterize novel molecules as analogues to these parameterized fragments via programs like Paramchem and MATCH[24–26]. This enables the large-sale parametrization of small molecules for both simulation of complexes and evaluation of docked poses which in turn enables models of higher precision to be applied in the field of docking-based drug development.

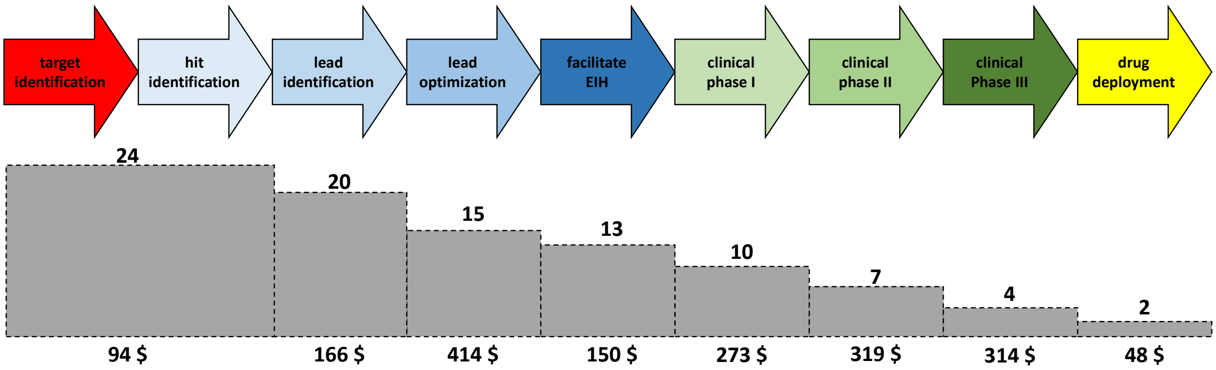


Figure 1: Schematic overview of the contemporary drug discovery process (top) and the number of ongoing projects (bottom gray upper) with the associated costs (bottom gray lower) in million US dollars[27].

1.3 Computer-aided drug discovery

Rational drug design describes the directed discovery and optimization of small organic molecules towards potency against a specific target with little side effects. The drug design process starts from initial target and hit identification (Figure 1). Subsequently, the hit molecule is further optimized into a lead structure of higher potency compared to the initial hit. In the next step, enhancement of properties of adsorption, distribution, metabolization, excretion and toxicity (ADMET) is emphasized. Molecules displaying high potency towards the target and simultaneously exhibit favorable metabolic properties are then studied in clinical trials, to be assessed for their applicability as a drug (Figure 1) which is then finally commissioned to the market.

In technical terms, the identification of small molecule inhibitors can be considered as an optimization and a combinatorial problem, dealing with the selection of a handful of active compounds from what is now considered to be a chemical space with a size that ranges between 10^{27} to 10^{100} synthetically available drug-like molecules[28–30], a number far too large for any chemist to synthesize and evaluate exhaustively. With the explosion of computational resources in the last decades, *in silico* modelling and evaluation of the interactions between a vast number of molecules and their designated target has become a feasibly and widely applied task called virtual screening. The aforementioned hit identification is the very basis of any virtual screening and describes the determination of a heterogeneous series of molecular scaffolds exhibiting homogeneous effects on the target protein. These effect are usually the displacement of the natural substrate/ligand of the target protein or other interference with the target function. There

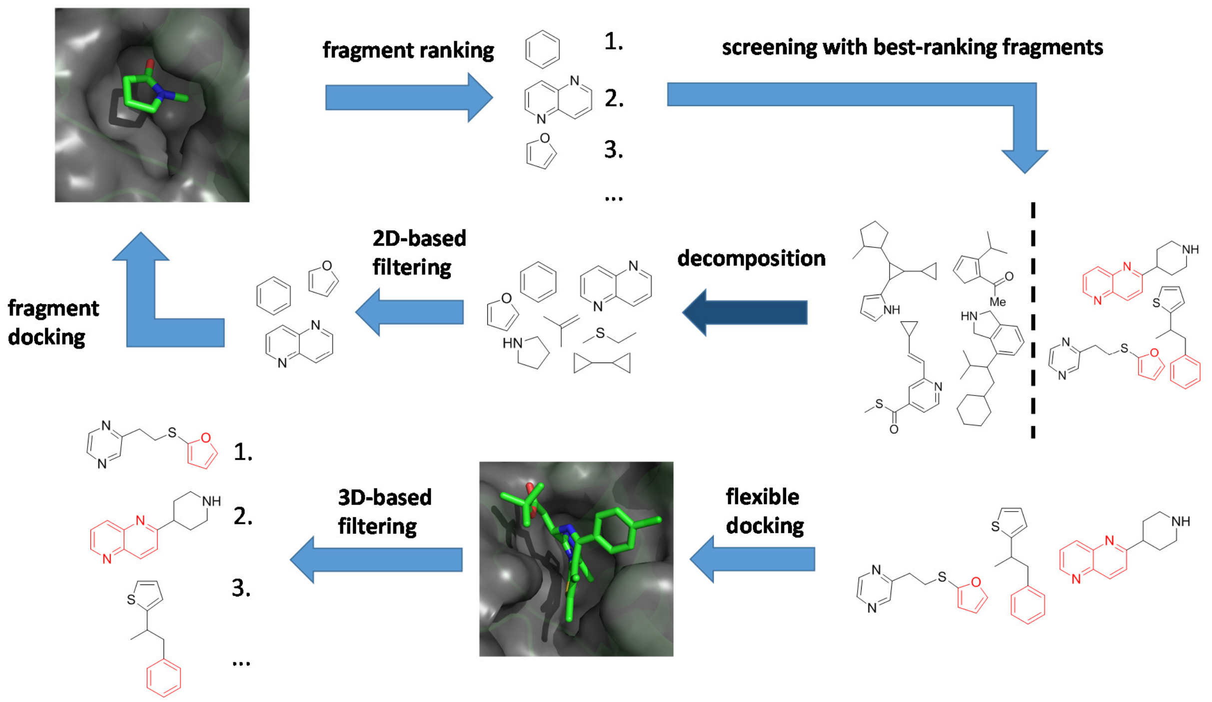


Figure 2: Overview of the ALTA procedure[31]. Beginning at the decomposition stage (dark blue arrow) a library of fragments is generated which is then filtered according to certain criteria like the Rule of Three for fragment-based design[32]. The fragments are then docked and looked up in larger molecules which contain them. The resulting parent molecules are then subjected to flexible docking and finally filtered to criteria resulting from the docking stage.

are a multitude of inhibition mechanisms by which this interference is achieved. Competitive inhibition takes place by direct displacement of the substrate via an inhibitor of higher potency. Allosteric inhibition, in which the inhibitor binds to a peripheral or secondary binding site which causes the displacement of the natural ligand in the active, site is a common form of non-competition-based inhibition. Additionally, there are uncompetitive inhibition mechanisms where the inhibitors achieves inhibition at the level of the protein-substrate-complex[33].

Once a promising scaffold is identified, its potency and optimizability towards more drug-like properties are investigated or created in what is considered lead identification. Metrics like the ligand efficiency (LE) and the ligand lipophilic efficiency (LLE) guide this selection process. Finally, a lead structure is further optimized into a drug candidate which is then ultimately assessed for its *in vivo* properties and goes into clinical trials (EIH, Entry Into Humans). If all clinical trials are passed, the compound is finally licensed and enters the market as a drug. The costs of the entire process are currently reported to exceed 800 million USD in 2004 towards 2.6 billion USD in 2014[27,34,35].

The virtual design of a ligand can incorporate the presence of structural information of the protein (structure-based design) or by predicting the binding properties of a novel molecule by inference from known properties of established ligands (ligand-based design). Lipinski's Rule of Five is probably the most widely applied ligand-based descriptor framing the upper and lower limits of xenobiotics absorption in humans[36]. Today, both methodologies are applied to ensure maximum success of the development at each step of the design process[37]. A particularly successful virtual screening scheme is fragment-based design[38]. Fragment-based design starts from small, weak-binding fragments which are subsequently extended into larger molecules. The way these fragments are generated can be different, depending on what the goals and restrictions of the drug design. Molecular fragmentation rules like RECAP have been devised to enable the chemical feasibility of the full molecules by ensuring that fragments satisfy retro-synthetic rules while other fragmentation approaches aim at the combinatorically largest set of fragments[39–41]. If the goal is *de novo* design in order to investigate chemical space not yet purchasable, fragments can be generated entirely by combinatorial enumeration of all possible structures[42]. The two main ways to extend fragments into larger molecules are growing and linking[43]. Growing extends the original fragments into other parts of the protein-ligand-interaction interface (now referred to as binding pocket) by ensuring that each addition to the fragment results in a larger structure with higher potency and specificity. Linking combines fragments which were placed in different parts of the binding pocket by a series of linker molecules[43]. The linking step is subject to a quality check similarly to the growing step, as only linkers providing favorable energy with the receptor are accepted into the molecule. A special case of fragment-based design is the so called anchor-based library tailoring approach (ALTA)[31]. The ALTA approach (Figure 2) exploits the fragment stage in order to reduce chemical space of commercially available compound libraries by identifying fragments with preferred binding properties. Depending on the amount of fragments simultaneously required in a larger molecule, the ALTA approach can be considered either a growing or linking scheme. Two prime advantages of this approach are the bypassing of the need for dedicated linking or growing software while ensuring that the resulting molecules are all chemically feasible. However, the prime disadvantage of this method is its inability to provide *de novo*-generated molecules as the chemical space which is screened is identical to the chemical space the fragments were generated from.

1.4 Structure of the thesis

The thesis is organized in the following manner: chapter 2 presents the experimental and computational studies performed on the cavitand system where a close correlation between simulation and experiments is highlighted. These studies resulted in respective highlighted publication. Chapter 3 deals with the published study of molecular dynamics simulations performed on CYP3A4 in complex with CBZ and the subsequent suggestion of mutants and their *in vitro* verification. Chapter 4 is a manuscript in preparation based on the virtual screening performed on SMARCA4 which resulted in the *in vitro* verified trans-glycoluril scaffold with an IC_{50} value of 20 μM. Chapter 5 is a regular thesis chapter not yet published on the identification of the putative site of interaction of iriomoteolide 3a and its target, monomeric actin as well as simulations of bound and unbound reidispongiolide A supporting its proposed mode of interaction with F-actin.

References

1. Van Gunsteren WF, Dolenc J, Mark AE. Molecular simulation as an aid to experimentalists. *Curr Opin Struct Biol*. 2008;18(2):149-153. doi:10.1016/j.sbi.2007.12.007.

2. Iavarone AT, Patriksson A, van der Spoel D, Parks JH. Fluorescence Probe of Trp-Cage Protein Conformation in Solution and in Gas Phase. *J Am Chem Soc*. 2007;129(21):6726-6735. doi:10.1021/ja065092s.

3. Gnanakaran S, Hochstrasser RM, García AE. Nature of structural inhomogeneities on folding a helix and their influence on spectral measurements. *Proc Natl Acad Sci U S A*. 2004;101(25):9229-9234. doi:10.1073/pnas.0402933101.

4. Gsponer J, Hopearuoho H, Cavalli A, Dobson CM, Vendruscolo M. Geometry, energetics, and dynamics of hydrogen bonds in proteins: structural information derived from NMR scalar couplings. *J Am Chem Soc*. 2006;128(47):15127-15135. doi:10.1021/ja0614722.

5. Peter C, Rueping M, Wörner HJ, Jaun B, Seebach D, van Gunsteren WF. Molecular Dynamics Simulations of Small Peptides: Can One Derive Conformational Preferences from ROESY Spectra? *Chem – Eur J*. 2003;9(23):5838-5849. doi:10.1002/chem.200305147.

6. Showalter SA, Brüschweiler R. Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field. *J Chem Theory Comput*. 2007;3(3):961-975. doi:10.1021/ct7000045.

7. Paci E, Gsponer J, Salvatella X, Vendruscolo M. Molecular Dynamics Studies of the Process of Amyloid Aggregation of Peptide Fragments of Transthyretin. *J Mol Biol*. 2004;340(3):555-569. doi:10.1016/j.jmb.2004.05.009.

8. Marrink SJ, Tieleman DP, Mark AE. Molecular Dynamics Simulation of the Kinetics of Spontaneous Micelle Formation. *J Phys Chem B*. 2000;104(51):12165-12173. doi:10.1021/jp001898h.

9. Huang D, Caflisch A. The free energy landscape of small molecule unbinding. *PLoS Comput Biol*. 2011;7(2):e1002002. doi:10.1371/journal.pcbi.1002002.

10. Hansen N, van Gunsteren WF. Practical Aspects of Free-Energy Calculations: A Review. *J Chem Theory Comput*. 2014;10(7):2632-2647. doi:10.1021/ct500161f.

11. Hou T, Wang J, Li Y, Wang W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model*. 2011;51(1):69-82. doi:10.1021/ci100275a.

12. Huang D, Caflisch A. Efficient Evaluation of Binding Free Energy Using Continuum Electrostatics Solvation. *J Med Chem*. 2004;47(23):5791-5797. doi:10.1021/jm049726m.

13. nobelprize.org. The Nobel Prize in Chemistry 2013, Nobel Media AB 2014. Web. 18 May 2015.http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/. Accessed May 18, 2015.

14. MacKerell AD, Brooks B, Brooks CL, et al. CHARMM: The Energy Function and Its Parameterization. In: *Encyclopedia of Computational Chemistry*. Vol John Wiley & Sons, Ltd; 2002. http://onlinelibrary.wiley.com/doi/10.1002/0470845015.cfa007/abstract. Accessed March 28, 2015.

15. Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: The biomolecular simulation program. *J Comput Chem*. 2009;30(10):1545-1614. doi:10.1002/jcc.21287.

16. Cornell WD, Cieplak P, Bayly CI, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc*. 1995;117(19):5179-5197. doi:10.1021/ja00124a002.

17. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc*. 1988;110(6):1657-1666. doi:10.1021/ja00214a001.

18. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*. 2004;25(13):1656-1676. doi:10.1002/jcc.20090.

19. Clark M, Cramer RD, Van Opdenbosch N. Validation of the general purpose tripos 5.2 force field. *J Comput Chem*. 1989;10(8):982-1012. doi:10.1002/jcc.540100804.

20. Molecular M, Field F, Halgren TA. Merck Molecular Force Field. 1996;17:490-519.

21. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc*. 1996;118(45):11225-11236. doi:10.1021/ja9621760.

22. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004;25(9):1157-1174. doi:10.1002/jcc.20035.

23. Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010;31(4):671-690. doi:10.1002/jcc.21367.

24. Yesselman JD, Price DJ, Knight JL, Brooks CL. MATCH: An atom-typing toolset for molecular mechanics force fields. *J Comput Chem*. 2012;33(2):189-202. doi:10.1002/jcc.21963.

25. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J Chem Inf Model*. 2012;52(12):3144-3154. doi:10.1021/ci300363c.

26. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J Chem Inf Model*. 2012;52(12):3155-3168. doi:10.1021/ci3003649.

27. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9(3):203-214. doi:10.1038/nrd3078.

28. Walters WP, Stahl MT, Murcko MA. Virtual screening—an overview. *Drug Discov Today*. 1998;3(4):160-178. doi:10.1016/S1359-6446(97)01163-X.

29. Ertl P. Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups. *J Chem Inf Comput Sci*. 2003;43(2):374-380. doi:10.1021/ci0255782.

30. Polishchuk PG, Madzhidov TI, Varnek A. Estimation of the size of drug-like chemical space based on GDB-17 data. *J Comput Aided Mol Des*. 2013;27(8):675-679. doi:10.1007/s10822-013-9672-4.

31. Kolb P, Kipouros CB, Huang D, Caflisch A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins*. 2008;73(1):11-18. doi:10.1002/prot.22028.

32. Congreve M, Carr R, Murray C, Jhoti H. A "Rule of Three" for fragment-based lead discovery? *Drug Discov Today*. 2003;8(19):876-877. doi:10.1016/S1359-6446(03)02831-9.

33. Cleland WW. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: II. Inhibition: Nomenclature and theory. *Biochim Biophys Acta BBA - Spec Sect Enzymol Subj*. 1963;67:173-187. doi:10.1016/0926-6569(63)90226-8.

34. Rawlins MD. Cutting the cost of drug development? *Nat Rev Drug Discov*. 2004;3(4):360-364. doi:10.1038/nrd1347.

35. Mullard A. New drugs cost US$2.6 billion to develop. *Nat Rev Drug Discov*. 2014;13(12):877-877. doi:10.1038/nrd4507.

36. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46(1-3):3-26.

37. Wilson GL, Lill MA. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med Chem*. 2011;3(6):735-750. doi:10.4155/fmc.11.18.

38. EVERTS S. PIECE BY PIECE. *Chem Eng News Arch*. 2008;86(29):15-23. doi:10.1021/cen-v086n029.p015.

39. Lewell XQ, Judd DB, Watson SP, Hann MM. RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci*. 1998;38(3):511-522.

40. Schuffenhauer A, Ertl P, Roggo S, Wetzel S, Koch M a, Waldmann H. The scaffold tree–visualization of the scaffold universe by hierarchical scaffold classification. *J Chem Inf Model*. 2007;47(1):47-58. doi:10.1021/ci600338x.

41. Kolb P, Caflisch A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J Med Chem*. 2006;49(25):7384-7392. doi:10.1021/jm060838i.

42. Schüller A, Hähnke V, Schneider G. SmiLib v2.0: A Java-Based Tool for Rapid Combinatorial Library Enumeration. *QSAR Comb Sci*. 2007;26(3):407-410. doi:10.1002/qsar.200630101.

43. Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov*. 2005;4(8):649-663. doi:10.1038/nrd1799.

Chapter 2

Experimental and Computational Study of BODIPY Dye-Labeled Cavitand Dynamics

Igor Pochorovski, Tim Knehans, Daniel Nettels, Astrid M. Müller, W. Bernd Schweizer, Amedeo Caflisch, Benjamin Schuler, and François Diederich

J. Am. Chem. Soc. 2014, 136, 2441–2449

J|A|C|S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article
pubs.acs.org/JACS

Experimental and Computational Study of BODIPY Dye-Labeled Cavitand Dynamics

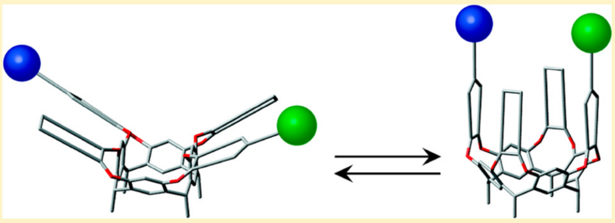
Igor Pochorovski,[†] Tim Knehans,[‡] Daniel Nettels,[‡] Astrid M. Müller,[§] W. Bernd Schweizer,[†] Amedeo Caflisch,*[,‡] Benjamin Schuler,*[,‡] and François Diederich*[,†]

[†]Laboratorium für Organische Chemie, ETH Zürich, Hönggerberg, HCI, 8093 Zürich, Switzerland

[‡]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[§]Beckman Institute, and Division of Chemistry and Chemical Engineering, California Institute of Technology, M/C 139-74, Pasadena, California 91125, United States

Supporting Information



ABSTRACT: Understanding the distance distribution and dynamics between moieties attached to the walls of a resorcin[4]arene cavitand, which is switchable between an expanded kite and a contracted vase form, might enable the use of this molecular system for the study of fundamental distance-dependent interactions. Toward this goal, a combined experimental and molecular dynamics (MD) simulation study on donor/acceptor borondipyrromethene (BODIPY) dye-labeled cavitands present in the vase and kite forms was performed. Direct comparison between anisotropy decays calculated from MD simulations with experimental fluorescence anisotropy data showed excellent agreement, indicating that the simulations provide an accurate representation of the dynamics of the system. Distance distributions between the BODIPY dyes were established by comparing time-resolved Förster resonance energy transfer experiments and MD simulations. Fluorescence intensity decay curves emulated on the basis of the MD trajectories showed good agreement with the experimental data, suggesting that the simulations present an accurate picture of the distance distributions and dynamics in this molecular system and provide an important tool for understanding the behavior of extended molecular systems and designing future applications.

INTRODUCTION

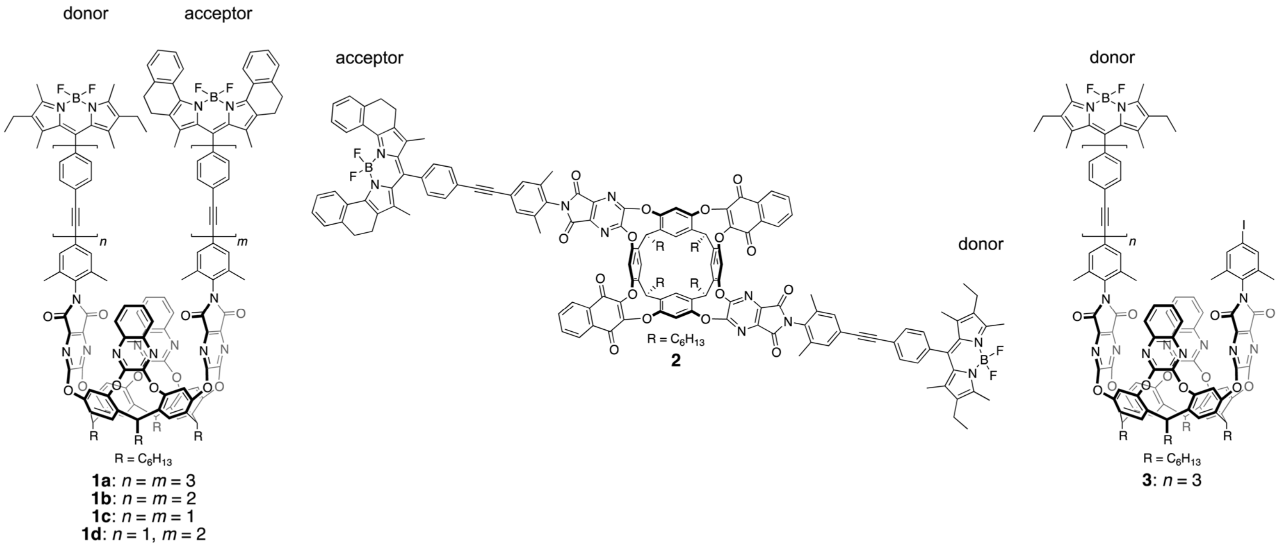
Resorcin[4]arene cavitands are a fascinating switching platform because of their ability to adopt two spatially well-defined conformations: an expanded kite and a contracted vase. Switching the conformational and binding properties of cavitands has been achieved with a variety of stimuli, such as changes in temperature,[1] pH,[2] metal ion concentration,[3] light irradiation,[4] solvent,[5] and redox state.[6] Besides employing conformational switching of resorcin[4]arene cavitands as a means to change their binding properties, the cavitand system could eventually be used as a platform to investigate fundamental interactions between objects attached to the cavitand's walls with respect to their variable distance. The development of such molecular machines able to controllably perform mechanical motions involving large spatial rearrangements is a long-standing goal.[7] The cavitand system could be used for this purpose if the distance distribution and dynamics between objects connected to the cavitand can be precisely determined in both the vase and kite conformations.

We had set out toward this goal by preparing a donor–acceptor borondipyrromethene (BODIPY) dye-substituted cavitand **1a** (Chart 1), present in the vase form, for Förster resonance energy transfer (FRET)[8] studies.[7c,9] Surprisingly, an unexpectedly low FRET efficiency was observed already in the vase form (ca. 66%), although an efficiency close to 100% was expected assuming a close dye–dye distance of ca. 1 nm. Such distance would prevail if the cavitand arms were oriented parallel to one another, i.e., the average opening angle of the cavitand walls would be 0°. The low FRET efficiency was attributed to either dynamic behavior of the cavitand or an unfavorable orientation of the transition dipole moments of the dyes.[7c,9] To gain more insights into this problem, cavitand **1a** was resynthesized together with the analogous cavitands with shorter phenylene–ethynylene linkers, **1b** and **1c**.[10] As the FRET efficiencies increased toward cavitands with shorter linkers (in the series **1a**, **1b**, **1c**), it was concluded that the two arms of the cavitands are separated by a certain average opening angle and are not aligned parallel. Thereby, an average opening angle of 16° was inferred to explain the observed FRET efficiencies.[10]

In this work, we sought to gain detailed insights into the distance distribution and dynamics of the cavitand system in both the vase and the kite conformations. Toward this goal, we prepared BODIPY dye-labeled quinone-based cavitand **2** (Chart 1) that is present in the kite form (designed based on

Received: October 16, 2013
Published: February 3, 2014

Chart 1. Structures of Vase Cavitands 1a−d, Kite Cavitand 2, and Donor-Only Substituted Reference Cavitand 3



1a: $n = m = 3$
1b: $n = m = 2$
1c: $n = m = 1$
1d: $n = 1$, $m = 2$

2

3: $n = 3$

what we learned about conformational properties of diquinone−diquinoxaline cavitands).[5,6] In addition, we expanded the BODIPY dye-labeled cavitand vase series **1a**−**c** by cavitand **1d** featuring phenylene−ethynylene linkers of different lengths. Together, systems **1a**−**d** and **2** embody the two conformational extremes of cavitands and are therefore ideally suited for the projected investigation. We investigated the BODIPY dye-labeled cavitands in a combined experimental and theoretical study consisting of time-resolved fluorescence spectroscopy and molecular dynamics (MD) simulations. BODIPY donor dye-substituted cavitand **3** shown in Chart 1 served as a reference compound for fluorescence studies. MD simulations complemented the experimental results and yielded theoretical dye−dye distance distributions. Emulation of fluorescence decay curves based on the distance distributions obtained by MD simulations allowed direct comparison of experimental and theoretical results. This work not only provides insights into the conformational dynamics of cavitands but also serves as a case study for the interplay between time-resolved fluorescence spectroscopy and MD simulations—a combination of methods that is often used to investigate the dynamics of biological macromolecules[11]—applied to a relatively simple artificial small molecule system.

■ RESULTS AND DISCUSSION

Synthesis and Characterization. The synthesis of cavitands **1a**−**c** and **3** had been recently reported.[10] The synthesis and characterization of the newly prepared cavitands **1d** and **2** is described in section 1 of the Supporting Information.

Cavitands **1a**−**d** and **2** differ in that **1a**−**d** possess two quinoxaline walls, while **2** is equipped with two quinone walls. This small structural difference has a dramatic effect on cavitand conformation:[5] cavitands **1a**−**c** are present in the vase form (^{1}H NMR methine proton shifts at 5.61 and 5.69 ppm),[1,12] while cavitand **2** adopts the kite form (methine protons at 3.69 and 4.35 ppm) in CDCl$_3$ solution.

^{19}F NMR spectroscopy employing a pulse sequence with a 30° flip angle enabled determination of the BODIPY dye donor/acceptor ratios in cavitands **1a**−**d** and **2** based on the integral ratios of the respective BF$_2$ units. In cavitand **1a** this ratio was 1.00/0.82, corresponding to a donor-only fraction of 18%. In cavitand **1b** the donor-only fraction was 11%. On the other hand, in cavitands **1c**, **1d**, and **2**, the donor-only fractions were below the sensitivity limit of the NMR measurements (∼2%). A possible explanation for higher ratios between donor/acceptor F atoms could be partial loss of the BF$_2$ units of the acceptor dyes during the course of cavitand syntheses, resulting in mixtures of fully labeled cavitands and cavitands lacking the BF$_2$ unit on the acceptor dyes.[13] Loss of the BF$_2$ unit in BODIPY dyes has precedence and was observed in high acidity media[14] or under strongly basic conditions,[15] and the corresponding products have been shown to be nonfluorescent.[14] Nevertheless, this finding is unexpected and has important implications for fluorescence studies, since the emission from this fraction of donor-only molecules needs to be taken into account for analysis of the fluorescence emission data.

X-ray Analysis. The X-ray structures of the precursors of cavitands **1a**−**d** and **2** and diiodocavitands **4**[16] and **5**[17] are shown in Figure 1. The solid-state conformational properties of the diiodocavitands are reflected by their solution-state properties: compound **4** crystallized in the vase form from (CH$_3$)$_2$CO/CH$_2$Cl$_2$, while compound **5** crystallized in the kite form from CDCl$_3$. While in cavitand **4** the I-bearing carbon atoms are placed at a distance of 0.86 nm to one another, in cavitand **5** the corresponding distance is 2.39 nm, which is almost 3 times larger.

Absorption and Steady-State Fluorescence Spectroscopy. The absorption and steady-state fluorescence spectra of cavitands **1a**−**d** and **2** are depicted in Figure 2 (top and bottom, respectively). The absorption spectra exhibit two main absorption bands corresponding to the donor dye moieties (λ_{max} = 529 nm) and the acceptor dye moieties (λ_{max} = 619 nm). The fluorescence spectra were recorded using an excitation wavelength of λ_{exc} = 490 nm; direct excitation of the acceptor is negligible at this wavelength.[18] The emission maxima at λ_{max} = 542 nm (I_{DA}) stem from the donor and at λ_{max} = 630 nm from the acceptor (I_{AD}) dye moieties. The intensity scale is referenced relative to the intensity maximum I_D of donor-only substituted cavitand **3**. The FRET efficiencies can be estimated according to eq 1

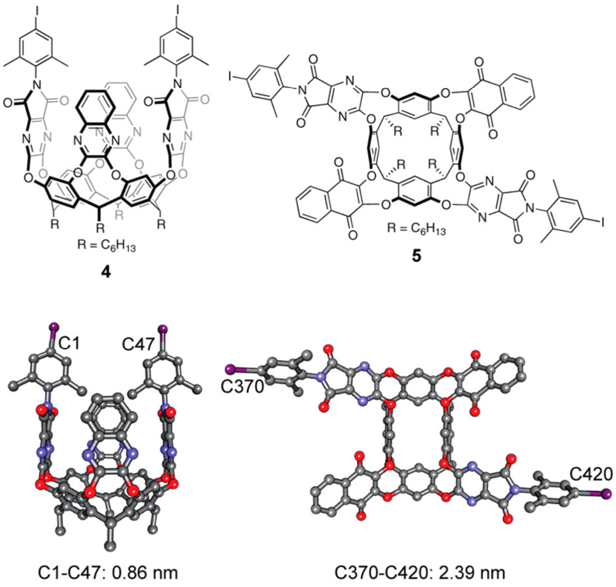


Figure 1. Molecular structures of 4^{16} (at 123 K) and 5 (at 100 K) in the crystals. Crystals of 4 and 5 were obtained by evaporation from $(CH_3)_2CO/CH_2Cl_2$ and $CDCl_3$, respectively. Solvent molecules, *n*-hexyl chains, and hydrogen atoms are omitted for clarity.

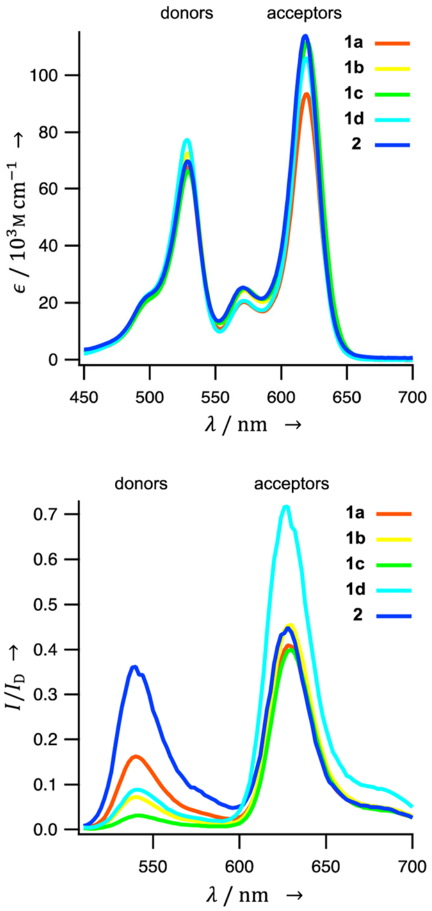


Figure 2. Absorption (top) and fluorescence emission (bottom, $c = 0.5 \times 10^{-7}$ M, $\lambda_{\rm exc}$ = 490 nm, $I_{\rm D}$ is the emission maximum of reference donor dye cavitand 3) spectra of cavitands **1a–d** and **2** in $CDCl_3$.

$$E = 1 - \frac{I_{\rm DA}}{I_{\rm D}} \tag{1}$$

Higher FRET efficiencies are expected for shorter dye–dye distances. Consequently, cavitand **2**, which is present in the kite conformation, exhibits the lowest FRET efficiency (64%) among the five cavitands. On the other hand, vase cavitands **1a–d** exhibit E values ranging from 84% to 97% with smaller values for cavitands with longer arms. While these efficiencies are higher than in cavitand **2**, suggesting shorter dye–dye distances, they are not consistent with the assumption of parallel oriented cavitand walls found in X-ray structures, in which case FRET efficiencies of over 99% would be expected. This trend was originally explained with the presence of a cavitand opening angle of ca. 16°.[10] In light of the above-mentioned finding based on NMR that samples of cavitands with longer arms possess significant donor-only labeled fractions (18% and 11% for cavitands **1a** and **1b**, respectively), the increasing donor fluorescence intensity for cavitands with longer arms could also be explained by residual fluorescence stemming from these contributions. This interpretation is supported by fluorescence lifetime measurements, where the contribution of doubly and singly labeled molecules can be resolved more easily.

Time-Resolved Fluorescence Spectroscopy. We recorded donor and acceptor fluorescence decay curves using time-correlated single-photon counting after pulsed excitation of the donor dye at $\lambda_{\rm exc}$ = 470 nm for each sample containing cavitands **1a–d** and **2** (Figure 3) at concentrations of ca. 10^{-6} M in $CHCl_3$. For a pure sample of donor–acceptor-labeled species with a single fixed distance, one expects to measure single-exponential donor fluorescence decays with a mean fluorescence lifetime $\tau_{\rm DA}$ reduced by a factor of $(1 - E)$ as compared to the mean donor fluorescence lifetime $\tau_{\rm D}$ obtained in the absence of the acceptor. Further, the acceptor decay curve should exhibit an initial rise (with rate constant $1/\tau_{\rm DA}$) due to the FRET-induced population of the acceptor excited state and a subsequent single-exponential decay with $1/\tau_{\rm A}$, where $\tau_{\rm A}$ is the mean fluorescence lifetime of the acceptor dye.[19] This behavior is clearly observed for the kite cavitand **2**. The corresponding donor decay curve was fitted with a single-exponential decay convolved with the instrument response function (IRF), yielding $\tau_{\rm DA}$ = 1.51 ns. This value results in a mean FRET efficiency of 64% (using $\tau_{\rm D}$ = 4.21 ns of reference cavitand **3** that is lacking the acceptor dye),[20] which is in excellent agreement with the steady-state fluorescence spectroscopy data. Note, however, that obtaining accurate distance information from this value requires the distance distribution and dynamics of the system to be taken into account.

The acceptor decay curves of vase cavitands **1a–d** are virtually identical and identical to the decay curve of the acceptor in **1a** directly excited with $\tau_{\rm exc}$ = 582 nm.[20] Their steep initial rise, which occurs within the response time of the instrument, shows that the FRET efficiencies in cavitands **1a–d** are near 100% and donor and acceptor dyes thus in very close proximity. The fluorescence emission of the donor is thus expected to be very weak, and the corresponding donor fluorescence lifetimes are very short. Note, however, that monitoring the sensitized acceptor emission has the advantage that only signal from molecules containing both a donor and an intact acceptor fluorophore is detected. In contrast, the fluorescence emission from the donor contains the signal of molecules lacking an active acceptor, whose presence we

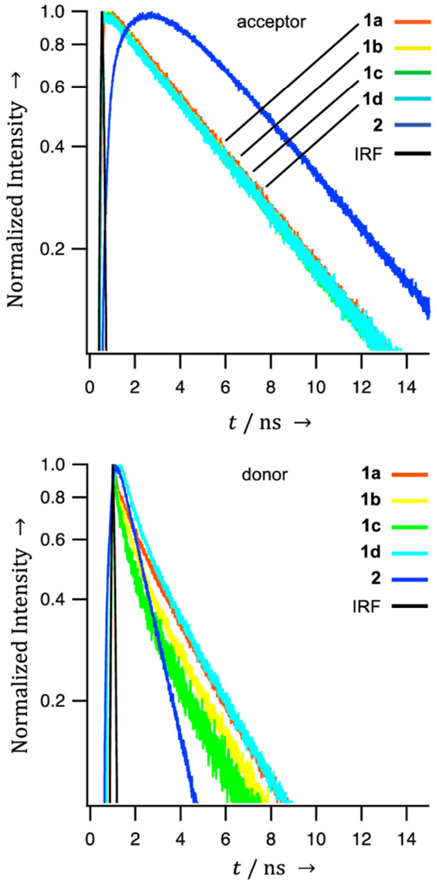


Figure 3. Fluorescence decay curves of the acceptor (top) and donor (bottom) moieties of cavitands **1a**–**d** and **2**. Measurements were performed under magic angle configuration, i.e., with the emission polarizer set to 54.7° with respect to the excitation polarization. IRF = Instrument response function.

quantified using NMR spectroscopy (see Synthesis and Characterization). The residual donor fluorescence intensity decays of cavitands **1a**–**d** (Figure 3b) are thus dominated by the emission from these donor-only molecules. As a result, they exhibit decay components up to the lifetime of the isolated donor fluorophore in the range of 4 ns[21] but also additional shorter lifetime components, presumably because of collisional quenching with the acceptor moiety lacking the BF_2 unit. Such multiexponential behavior is typical of BODIPY dye dimers or BODIPY dyes in confined environments (in proteins, lipids, micelles, or glasses).[22]

MD Simulations. To obtain a detailed molecular picture of the dye–dye distance distributions and dynamics underlying the fluorescence results, we performed MD simulations with explicit chloroform solvent for cavitands **1a**–**d** and **2**. For direct comparison to the experimental data and benchmarking of the MD results, we calculated observable properties such as fluorescence intensity and fluorescence anisotropy decay curves from the MD trajectories. Reports on MD simulations of supramolecular systems in explicit solvents are rare.[23] One reason for this is that early force fields were designed to simulate biomolecules in aqueous media. Recently, all-atom general force fields[24] and tools for automatic parameter assignment[24a,25] have been developed, as well as structure and topology files for various common organic solvents.[26] One goal of our study was therefore to utilize these developments and to investigate the applicability of MD simulations to molecular and supramolecular chemistry. The most common general force fields employed in contemporary MD literature are the General AMBER Force Field (GAFF)[24a] and the CHARMM General Force Field (CGenFF).[24b] For comparison, we included both force fields in our study.[27]

The most critical aspect in modeling cavitands **1a**–**d** and **2** are the phenylene–ethynylene linkers; small errors in their force field parameters would propagate to larger errors in BODIPY dye–dye distances. A common parameter in describing the stiffness of linker units is the persistence length L_p, whereby larger L_p values represent stiffer linkers. The persistence length of the phenylene–ethynylene unit had been recently experimentally determined on the basis of pulse electron paramagnetic resonance (EPR) spectroscopy on spin-labeled test systems.[28] We subjected these phenylene–ethynylene-containing systems (see the Supporting Information, section 6) to MD simulations with CGenFF and GAFF (explicit chloroform, 100 ns each) to investigate how well the two force fields reproduce the experimentally determined L_p value. Values of $L_p = 18.4 \pm 2.2$ nm for CGenFF and $L_p = 27.0 \pm 2.4$ nm for GAFF were obtained. Comparison to the experimental value of $L_p = 13.8 \pm 1.5$ nm[28] suggests that both force fields slightly overestimate the rigidity of the oligo(phenylene–ethynelene) linker, but especially the results from CGenFF provide reasonably good agreement.

The cavitands **1a**–**d** and **2** were subjected to MD simulations with the force fields CGenFF and GAFF (explicit chloroform, 500 ns each). The resulting BODIPY dye–dye distance histograms are illustrated in Figure 4. For the vase cavitands **1a**–**d**, both force fields yield histograms composed of sharp maxima at ca. 0.5 and 1.0 nm and broader distributions spanning from 0.5 to 1.9 nm in the case of the smallest cavitand **1c** and from 0.5 to 3.0 nm in the case of the largest cavitand **1a**. While the sharp maxima can be ascribed to arrangements where the BODIPY dyes are in direct contact, the broader distributions are due to motions of separated dyes.

On the other hand, the kite cavitand **2** yields single-distribution histograms spanning from 3.0 to 5.0 nm. The mean dye–dye distances $\langle d(\mathrm{B}\cdots\mathrm{B})\rangle$ obtained with the two force fields differ only marginally. The MD simulations slightly underestimate the average opening angle of cavitand **2**, as evidenced by comparing the C370–C420 distance (Figure 1) in the X-ray structure of cavitand **5** (2.39 nm) with the corresponding average C–C distance stemming from the CGenFF simulation of cavitand **2** (2.33 nm). If this discrepancy of 0.06 nm is propagated toward the BODIPY dyes, a small deviation of ca. 0.1 nm for the average simulated dye–dye distance can be expected.

Notably, there is no overlap between the dye–dye distance distributions of vase cavitand **1c** and kite cavitand **2**, which have the same linker length. Thus, switching between both cavitand conformations can entirely change the distance distribution profile of moieties attached to the cavitand walls.

Distance Autocorrelation from MD Data. To determine the time scale on which the distance dynamics of the BODIPY dye arms relative to each other takes place, we calculated the time-correlation functions[29]

$$g(\tau) = \frac{\langle R(t+\tau)R(t)\rangle_t}{\langle R(t)\rangle_t^2} \qquad (2)$$

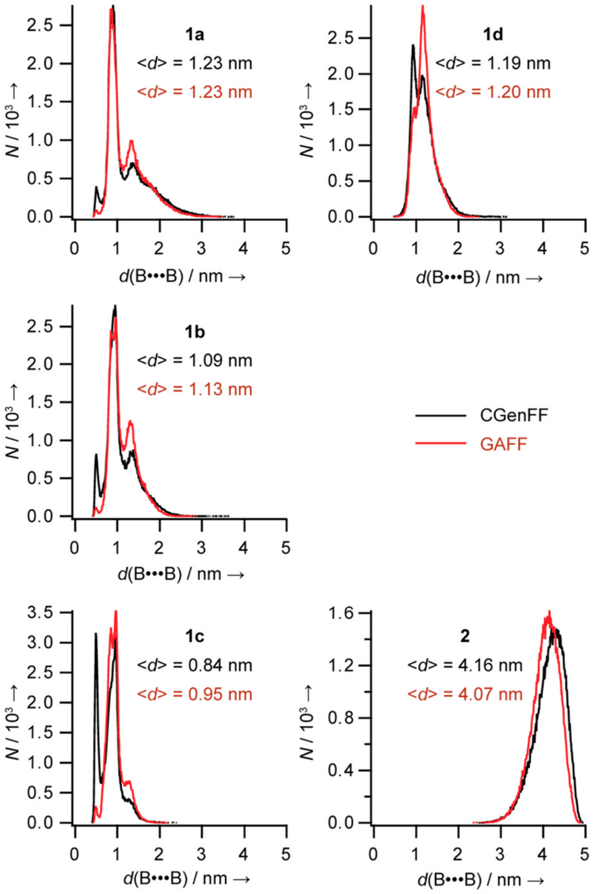


Figure 4. Histograms of distances between the B atoms of the donor and acceptor BODIPY dyes of cavitands **1a**−**d** and **2**, simulated with CGenFF (black) and GAFF (red).

from the CGenFF-simulated interdye distance data of cavitands **1a** and **2** (Figure 5); $R(t)$ is the dye−dye distance $d(\mathrm{B}\cdots\mathrm{B})$ at time t, $\langle\ldots\rangle_t$ denotes the time average over t, and τ is the lag time. The results show that the interdye distance dynamics of both cavitands occur on a subnanosecond time scale, with the dynamics of vase cavitand **1a** being slower than those of kite cavitand **2** by almost an order of magnitude. This finding is presumably due to a slow down of the distance dynamics in cavitand **1a** caused by dye−dye contacts in the closed conformation.

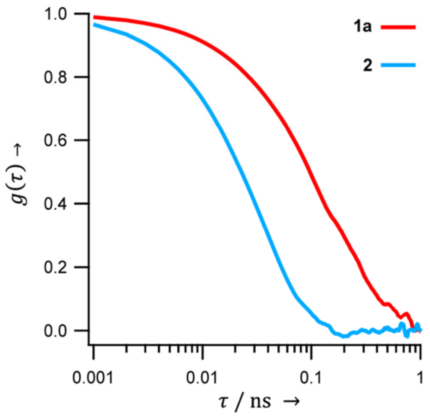


Figure 5. Autocorrelation functions of the interdye distances derived from CGenFF-MD simulations of cavitands **1a** and **2**.

Fluorescence Anisotropy. A stringent way of testing the accuracy of the time scales of dynamics in the simulations, including the viscosity of the solvent, is direct comparison to fluorescence anisotropy data. Anisotropy decays report on rotational diffusion times of the entire cavitands as well as on segmental rotation times of the BODIPY dye arms.[30] We measured the anisotropy decay of the acceptor dye of cavitand **1a** and compared it to the expected anisotropy decay as calculated from the corresponding MD simulations.

Fluorescence anisotropy decays were obtained by measuring the fluorescence intensity decays $I_{\mathrm{VV}}(t)$ and $I_{\mathrm{VH}}(t)$ observed after pulsed excitation of the acceptor dye (λ_{exc} = 582 nm); in both measurements the plane of the linear polarized excitation light was vertically oriented and the emission polarizer was set vertically for I_{VV} and horizontally for I_{VH} (Figure 6, top). We fitted the two curves globally with model curves $I_{\|}$ and $I_{\perp}$

$$I_{\|} = I_0(1 + 2r(t))e^{-t/\tau_{\mathrm{A}}} \qquad (3)$$

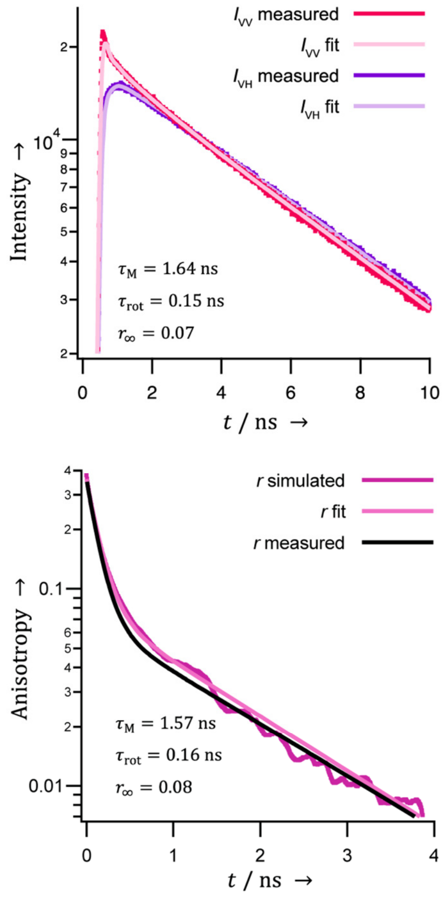


Figure 6. (Top) Acceptor fluorescence intensity decay curves of cavitand **1a** measured after acceptor excitation with λ_{exc} = 582 nm. I_{VV} was recorded with an emission polarization filter set to 0° and I_{VH} to 90° with respect to excitation polarization. (Bottom) Fluorescence anisotropy decay curves constructed from CGenFF-MD trajectory of cavitand **1a** using eq 6, fitted by the model described by eq 5, and the curve resulting from measured data inserted in eq 5.

$$I_\perp = I_0(1 - r(t))e^{-t/\tau_A} \tag{4}$$

τ_A is the mean fluorescence lifetime of the acceptor dye and was determined to be τ_A = 5.24 ns.[20] $I_\parallel$ and $I_\perp$ were convolved with the IRF to yield $I_{VV} = \mathrm{IRF}\otimes I_\parallel$ and $I_{VH} = G\cdot\mathrm{IRF}\otimes I_\perp$. The factor G describes the relative difference in detection efficiencies of vertical and horizontal polarized photons of the instrumentation.[30] For our instrument, we determined G = 1.1. The fluorescence anisotropy decay is defined as $r(t) = (I_\parallel - I_\perp)/(I_\parallel + 2I_\perp)$. We expected $r(t)$ to decay with two decay times: one corresponding to the overall rotational diffusion of the whole cavitand (τ_M), and the other corresponding to the rotational motion of the BODIPY dye relative to the cavitand (τ_{rot}). Hence, the anisotropy decay was described according to[31]

$$r(t) = ((r_0 - r_\infty)e^{-t/\tau_{rot}} + r_\infty)e^{-t/\tau_M} \tag{5}$$

where r_0 is the limiting anisotropy, which was fixed to 0.37,[22a] and r_∞ is the residual anisotropy. Fitting the intensity decays (Figure 6, top) yielded τ_M = 1.64 ns, τ_{rot} = 0.15 ns, and r_∞ = 0.07.

For direct comparison, we determined $r(t)$ also from the MD simulation of cavitand **1a**. The time trajectory of the normalized orientation vector $\vec{A}(t)$ of the acceptor dye was used to obtain the anisotropy decay according to[31]

$$r(t) = r_0\langle P_2(\vec{A}(t')\cdot\vec{A}(t' + t))\rangle_{t'} \tag{6}$$

Here, $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial and vector $\vec{A}$ is defined as indicated in Figure 7. We fitted the resulting anisotropy decay by eq 5 and obtained τ_M = 1.57 ns, τ_{rot} = 0.16 ns, and r_∞ = 0.08 (Figure 6, bottom). The excellent agreement between these values and the ones obtained from experimental anisotropy decay measurements provides good evidence that the MD simulations accurately capture the dynamics of the cavitands, including the effect of solvent and viscosity. The MD simulations should thus also be able to provide accurate insights in the combined effects of distance distributions and dynamics on the experimentally observed fluorescence intensity decays.

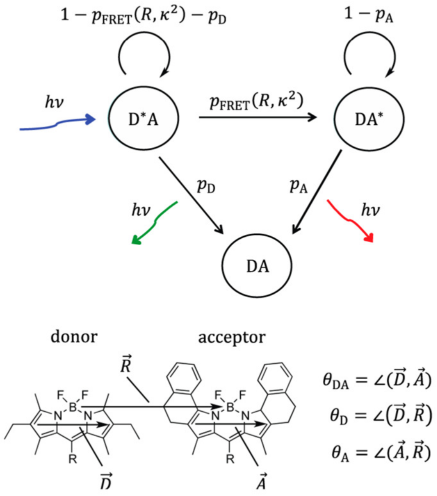


Figure 7. (Top) Markov chain model devised to emulate donor- and acceptor-fluorescence decay curves from simulated MD trajectories. (Bottom) Description of vectors and angles used for calculating κ^2 according to eq 12.

Fluorescence Decay Curves from MD Data. To enable a direct comparison of the simulations to the experimental data, we emulated donor fluorescence decay curves that would be expected on the basis of the CGenFF-simulated MD trajectories by taking into account the dye−dye distance and orientation time series (κ^2) and compared them with the measured decay curves. The decay curves were emulated using the Markov chain[32] model presented in Figure 7 (top), which was shown to be the best approach for calculating fluorescence observables.[11h−k] Starting from the donor excited state (D*A), the probabilities of FRET (p_{FRET}), donor emission (p_D), and acceptor emission (p_A) were calculated for every saved snapshot of the MD trajectory according to

$$p_{FRET} = (1 - e^{-(k_D + k_{FRET})\Delta t})\frac{k_{FRET}}{k_D + k_{FRET}} \tag{7}$$

$$p_D = (1 - e^{-(k_D + k_{FRET})\Delta t})\frac{k_D}{k_D + k_{FRET}} \tag{8}$$

$$p_A = 1 - e^{-k_A\Delta t} \tag{9}$$

with Δt = 4 ps being the time step at which coordinates were saved, k_{FRET} the rate constant of energy transfer, $k_d = 1/\tau_D$ = 0.24 ns^{-1}, and $k_A = 1/\tau_A$ = 0.19 ns^{-1} the respective fluorescence rate constants of the donor and acceptor dyes. The rate constant of energy transfer k_{FRET} was calculated from

$$k_{FRET} = k_D\left(\frac{R_0(\kappa^2)}{R}\right)^6 \tag{10}$$

with R being the dye−dye distance $d(\mathrm{B}\cdots\mathrm{B})$, and $R_0(\kappa^2)$ the dye−dye orientation-dependent Förster radius.[8] $R_0(\kappa^2)$ was calculated according to

$$R_0(\kappa^2) = R_0(2/3)\sqrt[6]{\frac{3}{2}\kappa^2} \tag{11}$$

The Förster radius R_0 (2/3) for κ^2 = 2/3 that is valid in case of freely rotating, isotropically averaged dyes was determined to be 4.91 nm. The orientation factor κ^2 was calculated according to[30]

$$\kappa^2 = (\cos\theta_{DA} - 3\cos\theta_D\cos\theta_A)^2 \tag{12}$$

with the angles defined by $\theta_{DA} = \angle(\vec{D},\vec{A})$,$\theta_D = \angle(\vec{D},\vec{R})$, and $\theta_A = \angle(\vec{A},\vec{R})$. The vectors $\vec{A}$ and $\vec{D}$ represent the emission transition dipole moment of the donor and the absorption transition dipole moment of the acceptor dye, respectively. $\vec{R}$ is the connection vector between the boron atoms of the dyes (Figure 7, bottom).

A total of 2500 different snapshots picked along regular intervals of 0.2 ns served as starting points for the Markov model. The model was started at state D*A and advanced by stepping through the trajectory. For every snapshot, transition probabilities according to eqs 7−9 were recalculated as functions of R and κ^2 (eqs 10−12). This process was stopped if photon emission from either the donor or the acceptor occurred or if the end of the trajectory was reached. The donor/acceptor photon counts at their respective emission times were summed up. To collect a statistically significant number of photon emission events, each of the 2500 runs was repeated 48 000 times to inject overall 120 millions of photons into each FRET emulation. The resulting decay curves were convolved with the respective IRFs to allow a direct comparison with experimental data.

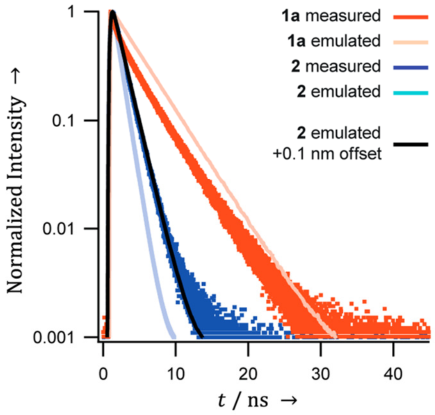


Figure 8. Measured and emulated donor fluorescence emission decay curves of vase cavitand **1a** (with an implemented donor-only fraction of 18%) and kite cavitand **2**.

The measured and emulated donor fluorescence emission decay curves of vase cavitand **1a** and kite cavitand **2** based on CGenFF trajectories are presented in Figure 8. When performing the emulation of the donor decay in cavitand **1a** without taking a donor-only-labeled fraction into account, an extremely rapid decay is observed that basically parallels the IRF.[33] The experimentally observed donor decay curve can only be reproduced if a donor-only labeled fraction is taken into account. Thus, 100% FRET efficiency is indeed expected for vase cavitands **1a**−**d** due to the close dye−dye distance and should not be obscured by a possibly suboptimal orientation of transition dipole moments. The deviation between the emulated and the measured curves stems from the fast decay component in the experimental curve that we attribute to dynamic collisional quenching between the fluorophores. The effect of quenching at low dye−dye distances is not described by Förster theory and is therefore not taken into account in the FRET emulation.[11i]

In contrast, the emulated donor decay curve of kite cavitand **2** is close to the experimental result, albeit with a slightly shorter average lifetime (1.01 ns) than the measured curve (1.51 ns), suggesting that the MD simulation underestimates the average dye−dye distance to some extent, most probably because of a slight deviation in the kite opening angle. Indeed, emulation of the donor fluorescence intensity decay curve on the basis of a simulated dye−dye distance distribution that is shifted by +0.1 nm yielded excellent overlap with the measured curve (Figure 8). This deviation in the average distance of ∼2% is in the same range as the difference between the results from the two force fields used (Figure 4) and illustrates the accuracy of MD simulations in reflecting the structural properties of molecular systems. With the information from the simulations, we can also test the accuracy of simple averaging regimes commonly used for analysis of FRET in dynamic systems.[11b,34] Since for cavitand **2** both the rotational correlation times of the dyes (Figure 6) and the distance dynamics between them (Figure 5) are much shorter than the fluorescence lifetimes, the system should be well approximated by the dynamic averaging regime, for which

$$\langle E\rangle = \frac{1}{1+\left(\int (R_0/r)^6 P(r)\,\mathrm{d}r\right)^{-1}} \tag{13}$$

Using the distance distribution of cavitand **2** from the simulations shifted by 0.1 nm as described above, we obtain an average transfer efficiency, $\langle E\rangle$, of 68%. The fair agreement with the experimental value of 64% indicates that dynamic averaging is a reasonable approximation.

■ CONCLUSIONS

The distance distribution and dynamics between moieties attached to resorcin[4]arene cavitand walls in both the kite and the vase conformations have been studied by a combination of experimental and theoretical methods on the basis of BODIPY dye-substituted vase cavitands **1a**−**d** and kite cavitand **2**. In kite cavitand **2**, featuring one phenylene−ethynylene linker unit per dye, the dye moieties adopt an average distance of ca. 4.2 nm according to MD simulations (CGenFF). Time-resolved fluorescence spectroscopy revealed a FRET efficiency of 64%. In the case of vase cavitands **1a**−**d**, steady-state and time-resolved fluorescence spectroscopy showed increasing amounts of donor fluorescence for cavitands with longer arms. In early studies, this trend was explained by increasing dye−dye distances for cavitands with longer arms, which would be consistent with the presence of a nonzero cavitand opening angle. In the current study, we found according to MD simulations (CGenFF) that while the average dye−dye distance does indeed increase from 0.8 nm in the shortest cavitand **1c** to 1.2 nm in the longest cavitand **1a**, these distances are still in a range that should yield FRET efficiencies of 100%. Instead, the increasing donor fluorescence for cavitands with longer arms could be explained by the previously undetected presence of cavitand fractions with inactive acceptor components lacking the BF_2 unit. Emulation of the fluorescence intensity decay curve of cavitand **1a** from simulated MD trajectories confirmed that the experimentally observed decay curve can only be reproduced when a donor-only-labeled cavitand fraction is taken into account.

The dynamics of vase cavitand **1a** was investigated by fluorescence anisotropy measurements. Evaluation of both measured and simulated fluorescence anisotropy decay curves showed that the rotation time of a BODIPY dye arm in cavitand **1a** is ca. 0.15 ns. Autocorrelation analysis of the dye−dye distance time series revealed that the distance dynamics of the BODIPY dye arms takes place on the subnanosecond time scale. Most importantly, the fluorescence lifetime decays calculated based on the MD simulations of cavitand **2** are in excellent agreement with the experimental data, showing that the simulations present an accurate picture of the distance distribution and dynamics in this molecular system. With these results, the cavitand system can now be used as a platform to investigate fundamental, distance-dependent interactions between objects attached to the cavitand's walls.

In the case of extended molecular and supramolecular systems with pronounced flexibility and therefore broad intramolecular distance distributions, the combination of time-resolved FRET with simulations is essential to quantify the underlying dynamics because the experimental observables depend on both the shape of the distance distribution and the time scale of the dynamics. In this work, we have in-depth structurally characterized the two conformational states of resorcin[4]arene cavitands, which opens the possibility to utilize the cavitand system for studying intraspace interactions ranging from 1 to 7 nm (adjustable via the linker length). Furthermore, the good agreement between MD simulations and experimental data should further encourage application of MD simulations to molecular chemical systems. Progress in force-field development now allows such simulations to be performed with high accuracy. MD simulations could serve as a testing tool for envisioned concepts and thereby guide the synthetic chemist toward successful implementation of his/her ideas.

Although the dyes used in this study are not suitable for single-molecule fluorescence detection, an approach analogous to the one demonstrated here could be employed for single-molecule studies. Such experiments could further enhance the resolution of structural and dynamic heterogeneities in molecular systems.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Synthetic procedures, characterization data, X-ray data, fluorescence data, MD methods, NMR spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

caflisch@bioc.uzh.ch
schuler@bioc.uzh.ch
diederich@org.chem.ethz.ch

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the Swiss National Science Foundation (SNF). I.P. acknowledges the receipt of a fellowship from the Fonds der Chemischen Industrie. Research was partly carried out in the Laser Resource Center of the Beckman Institute of the California Institute of Technology and supported by the Arnold and Mabel Beckman Foundation. We thank Prof. Dr. Gunnar Jeschke, Dr. Marc-Olivier Ebert, and Dr. Andreas Vitalis for helpful discussions, Dimitry Kotlyar for help with the Table of Contents artwork, and Dr. Kenno Vanommeslaeghe for providing parametrization of the cavitands via the Paramchem engine.

■ REFERENCES

(1) (a) Moran, J. R.; Ericson, J. L.; Dalcanale, E.; Bryant, J. A.; Knobler, C. B.; Cram, D. J. *J. Am. Chem. Soc.* **1991**, *113*, 5707. (b) Moran, J. R.; Karbach, S.; Cram, D. J. *J. Am. Chem. Soc.* **1982**, *104*, 5826.

(2) Skinner, P. J.; Cheetham, A. G.; Beeby, A.; Gramlich, V.; Diederich, F. *Helv. Chim. Acta* **2001**, *84*, 2146.

(3) (a) Frei, M.; Marotti, F.; Diederich, F. *Chem. Commun.* **2004**, 1362. (b) Durola, F.; Rebek, J., Jr. *Angew. Chem., Int. Ed.* **2010**, *49*, 3189.

(4) (a) Berryman, O. B.; Sather, A. C.; Lledo, A.; Rebek, J. *Angew. Chem., Int. Ed.* **2011**, *50*, 9400. (b) Berryman, O. B.; Sather, A. C.; Rebek, J., Jr. *Chem. Commun.* **2010**, *47*, 656.

(5) Pochorovski, I.; Boudon, C.; Gisselbrecht, J.-P.; Ebert, M.-O.; Schweizer, W. B.; Diederich, F. *Angew. Chem., Int. Ed.* **2012**, *51*, 262.

(6) Pochorovski, I.; Ebert, M.-O.; Gisselbrecht, J.-P.; Boudon, C.; Schweizer, W. B.; Diederich, F. *J. Am. Chem. Soc.* **2012**, *134*, 14702.

(7) (a) Barboiu, M.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5201. (b) Dietrich-Buchecker, C. O.; Jimenez-Molero, M. C.; Sartor, V.; Sauvage, J. P. *Pure Appl. Chem.* **2003**, *75*, 1383. (c) Azov, V. A.; Schlegel, A.; Diederich, F. *Angew. Chem., Int. Ed.* **2005**, *44*, 4635. (d) Azov, V. A.; Beeby, A.; Cacciarini, M.; Cheetham, A. G.; Diederich, F.; Frei, M.; Gimzewski, J. K.; Gramlich, V.; Hecht, B.; Jaun, B.; Latychevskaia, T.; Lieb, A.; Lill, Y.; Marotti, F.; Schlegel, A.; Schlittler, R. R.; Skinner, P. J.; Seiler, P.; Yamakoshi, Y. *Adv. Funct. Mater.* **2006**, *16*, 147. (e) Wu, J.; Leung, K. C.-F.; Benítez, D.; Han, J.-Y.; Cantrill, S. J.; Fang, L.; Stoddart, J. F. *Angew. Chem., Int. Ed.* **2008**, *47*, 7470.

(8) Förster, T. *Ann. Phys. (Berlin, Ger.)* **1948**, *2*, 55.

(9) Azov, V. A.; Schlegel, A.; Diederich, F. *Bull. Chem. Soc. Jpn.* **2006**, *79*, 1926.

(10) Pochorovski, I.; Breiten, B.; Schweizer, W. B.; Diederich, F. *Chem.—Eur. J.* **2010**, *16*, 12590.

(11) (a) Gustiananda, M.; Liggins, J. R.; Cummins, P. L.; Gready, J. E. *Biophys. J.* **2004**, *86*, 2467. (b) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754. (c) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1528. (d) Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18964. (e) Corry, B.; Jayatilaka, D. *Biophys. J.* **2008**, *95*, 2711. (f) Unruh, J. R.; Kuczera, K.; Johnson, C. K. *J. Phys. Chem. B* **2009**, *113*, 14381. (g) Schuetz, P.; Wuttke, R.; Schuler, B.; Caflisch, A. *J. Phys. Chem. B* **2010**, *114*, 15227. (h) Liao, J.-M.; Wang, Y.-T.; Chen, C.-L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10364. (i) Hoefling, M.; Lima, N.; Haenni, D.; Seidel, C. A. M.; Schuler, B.; Grubmüller, H. *PLoS ONE* **2011**, *6*, e19791. (j) Speelman, A. L.; Muñoz-Losa, A.; Hinkle, K. L.; VanBeek, D. B.; Mennucci, B.; Krueger, B. P. *J. Phys. Chem. A* **2011**, *115*,, 3997;. (k) Hoefling, M.; Grubmüller, H. *Comput. Phys. Commun.* **2013**, *184*, 841.

(12) Azov, V. A.; Jaun, B.; Diederich, F. *Helv. Chim. Acta* **2004**, *87*, 449.

(13) A more detailed discussion on the instability of the acceptor dye is provided in section 1.4 of the Supporting Information.

(14) (a) Liras, M.; Prieto, J. B.; Pintado-Sierra, M.; Arbeloa, F. L.; García-Moreno, I.; Costela, Á.; Infantes, L.; Sastre, R.; Amat-Guerri, F. *Org. Lett.* **2007**, *9*, 4183. (b) Bañuelos, J.; López Arbeloa, F.; Arbeloa, T.; Salleres, S.; Vilas, J. L.; Amat-Guerri, F.; Liras, M.; López Arbeloa, I. *J. Fluoresc.* **2008**, *18*, 899.

(15) (a) Crawford, S. M.; Thompson, A. *Org. Lett.* **2010**, *12*, 1424. (b) Smithen, D. A.; Baker, A. E. G.; Offman, M.; Crawford, S. M.; Cameron, T. S.; Thompson, A. *J. Org. Chem.* **2012**, *77*, 3439.

(16) Published in ref 10.

(17) For details, see section 2 of the Supporting Information.

(18) There are no indications that electron transfer, neither between the two dyes nor between the dyes and the quinone walls, significantly affects the transfer dynamics. For a more detailed discussion, see section 3.2 of the Supporting Information.

(19) Meer, B. W. V. d.; Coker, G.; Chen, S.-Y. S. *Resonance Energy Transfer: Theory and Data*; VCH Publishers, Inc.: New York, 1994.

(20) For details, see Figure S5 in section 3.1 of the Supporting Information.

(21) An analysis of the decay curves is provided in section 3.4 of the Supporting Information.

(22) (a) Karolin, J.; Johansson, L. B.-Å.; Strandberg, L.; Ny, T. *J. Am. Chem. Soc.* **1994**, *116*, 7801. (b) Bergström, F.; Hägglöf, P.; Karolin, J.; Ny, T.; Johansson, L. B.-Å. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12477. (c) Bergström, F.; Mikhalyov, I.; Hägglöf, P.; Wortmann, R.; Ny, T.; Johansson, L. B.-Å. *J. Am. Chem. Soc.* **2002**, *124*, 196. (d) Mikhalyov, I.; Gretskaya, N.; Bergström, F.; Johansson, L. B.-Å. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5663. (e) Tleugabulova, D.; Zhang, Z.; Brennan, J. D. *J. Phys. Chem. B* **2002**, *106*, 13133.

(23) (a) Fischer, S.; Grootenhuis, P. D. J.; Groenen, L. C.; van Hoorn, W. P.; van Veggel, F. C. J. M.; Reinhoudt, D. N.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 1611. (b) den Otter, W. K.; Briels, W. J. *J. Am. Chem. Soc.* **1998**, *120*, 13167. (c) Tolpekina, T. V.; den Otter, W. K.; Briels, W. J. *J. Phys. Chem. B* **2003**, *107*, 14476. (d) Casanovas, J.; Zanuy, D.; Aleman, C. *Angew. Chem., Int. Ed.* **2006**, *45*, 1103. (e) Ewell, J.; Gibb, B. C.; Rick, S. W. *J. Phys. Chem. B* **2008**, *112*, 10272. (f) Javor, S.; Rebek, J. *J. Am. Chem. Soc.* **2011**, *133*, 17473. (g) Cézard, C.; Trivelli, X.; Aubry, F.; Djedaïni-Pilard, F.; Dupradeau, F.-Y. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15103. (h) Zhang, H.; Tan, T.;

Feng, W.; van der Spoel, D. *J. Phys. Chem. B* **2012**, *116*, 12684. (i) Zheng, X.; Wang, D.; Shuai, Z.; Zhang, X. *J. Phys. Chem. B* **2012**, *116*, 823. (j) Simona, F.; Nussbaumer, A. L.; Häner, R.; Cascella, M. *J. Phys. Chem. B* **2013**, *117*, 2576.

(24) (a) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157. (b) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2009**, *31*, 672.

(25) (a) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247. (b) Vanommeslaeghe, K.; Mackerell, A. D., Jr. *J. Chem. Inf. Model.* **2012**, *52*, 3144. (c) Vanommeslaeghe, K.; Raman, E. P.; Mackerell, A. D., Jr. *J. Chem. Inf. Model.* **2012**, *52*, 3155.

(26) (a) van der Spoel, D.; van Maaren, P. J.; Caleman, C. *Bioinformatics* **2012**, *28*, 752. (b) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. *J. Chem. Theory Comput.* **2012**, *8*, 61.

(27) For details, see section 4 of the Supporting Information.

(28) (a) Godt, A.; Schulte, M.; Zimmermann, H.; Jeschke, G. *Angew. Chem., Int. Ed.* **2006**, *45*, 7560. (b) Jeschke, G.; Sajid, M.; Schulte, M.; Ramezanian, N.; Volkov, A.; Zimmermann, H.; Godt, A. *J. Am. Chem. Soc.* **2010**, *132*, 10107.

(29) Dunn, P. F. *Measurement and Data Analysis for Engineering and Science*; McGraw-Hill: New York, 2005.

(30) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd ed.; Kluwer Academic/Plenum Publishers: New York, 1999.

(31) Lipari, G.; Szabo, A. *Biophys. J.* **1980**, *30*, 489.

(32) Norris, J. R. *Markov Chains*; University of Cambridge: Cambridge, 1998.

(33) See Figure S9 in section 5 of the Supporting Information.

(34) (a) Schuler, B.; Müller-Späth, S.; Soranno, A.; Nettels, D. *Methods Mol. Biol.* **2012**, *896*, 21. (b) Wozniak, A. K.; Schröder, G. F.; Grubmüller, H.; Seidel, C. A. M.; Oesterhelt, F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18337.

Journal of the American Chemical Society

Supporting Information

Experimental and Computational Study of BODIPY Dye-Labeled Cavitand Dynamics

Igor Pochorovski, Tim Knehans, Daniel Nettels, Astrid M. Müller, W. Bernd Schweizer, Amedeo Caflisch, Benjamin Schuler, François Diederich

Table of Contents

S3

1 Synthesis Section

1.1 Synthetic Schemes

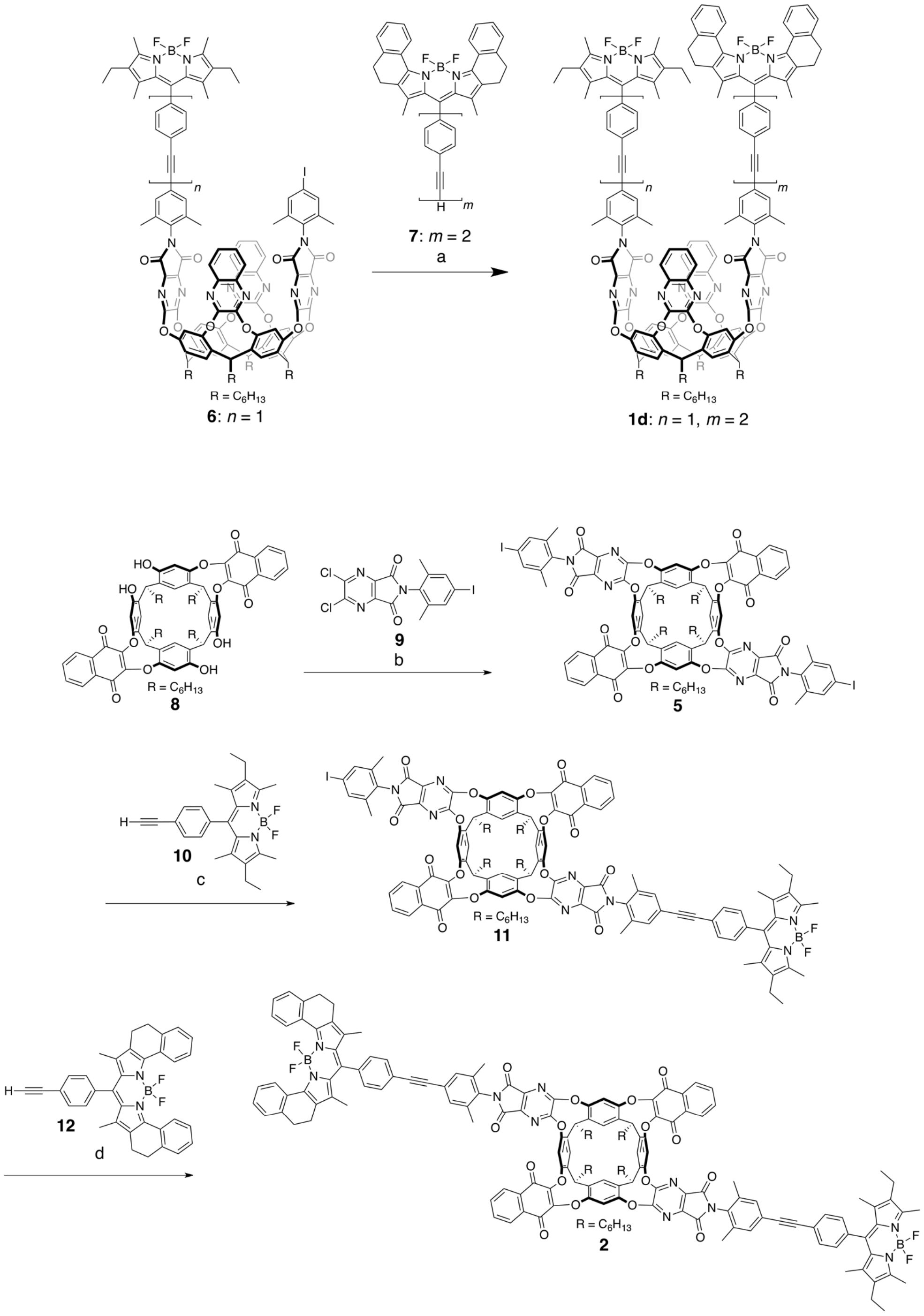


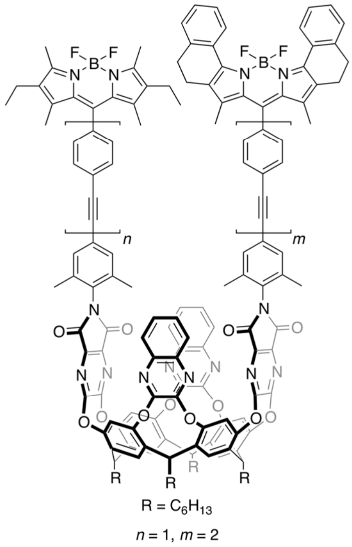
Figure S1. Synthesis of BODIPY dye-substituted cavitands **1d** and **2**. a) [Pd(PPh$_3$)$_4$}], CuI, *i*-Pr$_2$NEt, THF, 35 °C, 2 d; 37%. b) Cs$_2$CO$_3$, THF, 70 °C, 24 h; 77%. c) [Pd(PPh$_3$)$_4$}], CuI, *i*-Pr$_2$NEt, THF, 25 °C, 3 d; 17%. d) [Pd(PPh$_3$)$_4$}], CuI, *i*-Pr$_2$NEt, THF, 35 °C, 3 d; 11%.

1.2 Materials and General Methods

All chemicals were purchased as reagent grade and used without further purification. When stated, solvents were degassed by bubbling Ar through the solution for 30 min. Flash chromatography (FC) was performed using SiO_2-60 (230–400 mesh ASTM, 0.040–0.063 mm; Fluka) or SiO_2-F60 (0.040–0.063 mm, 60 Å, Silicycle). Preparative recycling gel permeation chromatography (GPC) was run on a Japan Analytical Industries LC-9101 preparative recycling HPLC apparatus using HPLC-grade $CHCl_3$ as the mobile phase. Melting points were measured on a Büchi B-540 melting-point apparatus in open capillaries. ^{1}H NMR, ^{13}C NMR, and ^{19}F NMR spectra were recorded on a Bruker DRX 400 or Bruker AV 400 spectrometer at 298 K. Residual solvent peaks were used as internal references. ATR Infrared spectra (IR) were recorded on a Varian 800 FT-IR spectrometer. Selected absorption bands are reported in wavenumbers (cm^{-1}). Mass spectrometry was performed by the MS-service at ETH Zürich. High-resolution electron impact mass spectra were measured on a Waters Micromass AutoSpec Ultima spectrometer. High-resolution matrix-assisted laser-desorption-ionization mass spectra were measured on a Varion Ionspec Ultima MALDI-FTICR mass spectrometer using 3-hydroxypyridine-2-carboxylic acid (3-HPA) as matrix or on Bruker Daltonics Ultraflex II MALDI-TOF mass spectrometer using (2-[(2E)-3-(4-t-butylphenyl)-2-methylprop-2-enylidene]malononitril) (DCTB) as matrix. High-resolution electro-spray-ionisation mass spectra were measured on a Bruker Daltonics maXis spectrometer. UV/Vis spectroscopy was carried out with a Varian Cary 500 Scan spectrophotometer. Steady state fluorescence spectroscopy was carried out with an Instruments S. A. Fluorolog-3 spectrofluorimeter. Both UV/Vis and fluorescence experiments were carried out in standard 3.5 mL quartz cells (4 optical windows for UV/Vis, 2 optical windows for fluorescence) with 10 mm path length.

1.3 Synthetic Procedures

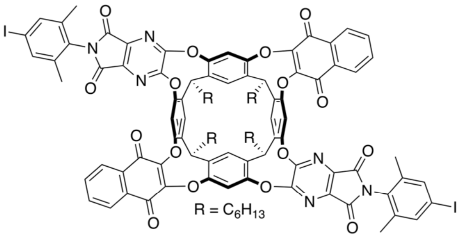
Compounds **6**[1], **7**[1-2], **8**[3], **9**[1-2], **10**[1-2], and **12**[1-2] were prepared according to literature procedures.



Cavitand 1d

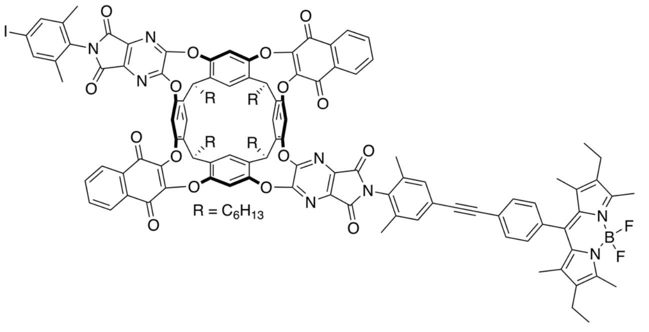
[Pd(PPh$_3$)$_4$] (19 mg, 0.016 mmol) and CuI (3 mg, 0.016 mmol) were added to a degassed solution of **6** (76 mg, 0.121 mmol), **7** (0.17 g, 0.081 mmol), and DIPEA (0.28 mL, 1.6 mmol) in THF (5 mL) and CHCl$_3$ (5 mL). The mixture was stirred for 2 d at 35 °C, after which the solvent was evaporated. FC (SiO$_2$; CH$_2$Cl$_2$ → CH$_2$Cl$_2$/EtOAc 98:2) and recycling GPC (Jaigel-2H; CHCl$_3$) afforded **1d** (78 mg, 37 %) as a blue solid. ^{1}H NMR (400 MHz, CDCl$_3$): δ = 0.90 – 1.05 (m, 18H), 1.27 – 1.64 (m, 50H), 2.16 – 2.41 (m, 19H), 2.51 – 2.63 (m, 10H), 2.83 – 2.99 (m, 4H), 5.63 (t, J = 8.1, 2H), 5.71 (t, J = 8.1, 2H), 7.26 – 7.64 (m, 26H), 7.70 (d, J = 8.1, 1H), 7.75 (d, J = 8.1, 2H), 7.85 – 7.95 (m, 4H), 8.26 (s, 2H), 8.26 (s, 2H), 8.82 ppm (d, J= 8.0, 2H); ^{13}C NMR (101 MHz, CDCl$_3$): δ = 11.96, 12.33, 12.56, 14.04, 14.60, 17.11, 18.05, 18.06, 18.08, 18.10, 20.49, 22.65, 27.93, 27.96, 29.33, 29.35, 30.53, 31.85, 31.86, 32.16, 32.70, 34.20, 34.30, 89.55, 90.20, 90.25, 90.37, 90.43, 90.97, 118.79, 122.99, 123.24, 123.51, 123.77, 125.02, 125.11, 127.32, 128.05, 128.39, 128.73, 129.02, 129.09, 129.14, 129.36, 129.44, 130.57, 131.21, 131.34, 131.66, 131.72, 131.81, 132.17, 132.29, 132.34, 132.36, 132.43, 133.03, 133.75, 135.70, 135.84, 136.12, 136.13, 136.34, 136.54, 136.86, 137.40, 137.51, 138.14, 138.76, 139.12, 139.82, 140.69, 141.48, 150.95, 152.04, 152.25, 153.09, 154.16, 158.85, 158.87, 161.43 ppm; ^{19}F NMR

(376 MHz, $CDCl_3$): $\delta = -145.63$ (q, $J = 31.7$, 2F), -134.96 (q, $J = 33.2$, 2F); IR (ATR): $\tilde{\nu} =$ 2926 (w), 2857 (w), 1741 (m), 1525 (m), 1480 (m), 1442 (w), 1411 (m), 1362 (m), 1326 (s), 1275 (m), 1231 (m), 1191 (s), 1157 (s), 1114 (m), 1082 (s), 978 (m), 898 (m), 837 (m), 792 (w), 761 (s), 709 (m), 688 (m), 626 (w); UV/Vis ($CHCl_3$): λ_{max} (ε) = 530 (69000), 620 nm (110000); HR-MALDI-MS (3-HPA): *m*/*z* (%): 2581.1280 (100, $[M-F]^+$, calcd for $C_{164}H_{144}B_2F_3N_{14}O_{12}^+$: 2581.1260).



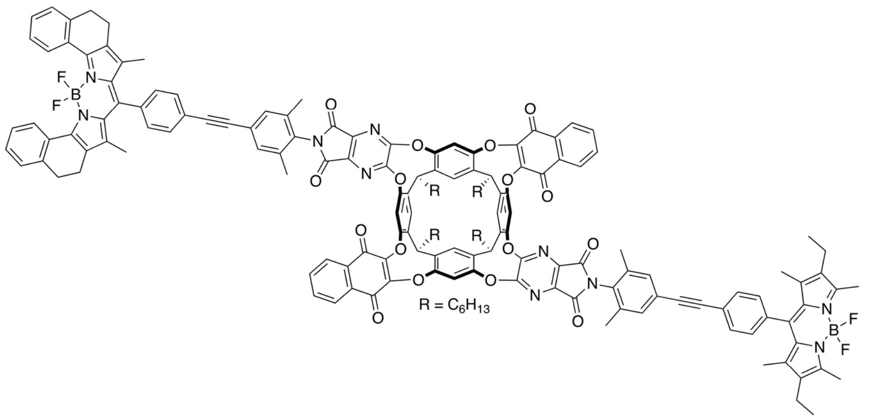
Cavitand 5

Cs_2CO_3 (2.17 g, 6.67 mmol) was added to a solution of tetrol **8** (1.8 g, 1.6 mmol) and iodoimide **9** (1.49 g, 3.34 mmol) in THF (100 mL). The mixture was stirred for 24 h at 70 °C, filtered over silica, and the solvent was evaporated. FC (SiO_2; $CH_2Cl_2 \rightarrow CH_2Cl_2$/EtOAc 95:5) afforded **5** (2.3 g, 77 %) as a yellow solid. $R_f = 0.48$ (SiO_2; CH_2Cl_2/EtOAc 98:2); m.p. > 320 °C (decomp); ^1H NMR (400 MHz, $CDCl_3$): δ = 0.79 – 0.96 (m, 12H), 1.21 – 1.41 (m, 32H), 2.01 – 2.15 (m, 14H), 2.17 (s, 6H), 3.68 (t, $J = 7.5$, 2H), 4.37 (t, $J = 7.5$, 2H), 6.15 – 7.23 (m, 4H), 7.42 (br s, 4H), 7.57 (s, 2H), 7.60 (s, 2H), 7.65 – 7.80 (m, 4H), 8.02 – 8.26 ppm (m, 4H); ^{13}C NMR (101 MHz, $CDCl_3$): δ = 13.97, 13.99, 17.74, 17.80, 22.52, 22.58, 26.96, 27.05, 29.04, 29.12, 31.49, 31.60, 31.66, 35.55, 37.70, 96.21, 126.76, 128.88, 130.30, 134.35, 137.60, 137.69, 138.84, 139.05, 151.77, 161.78 ppm; IR (ATR): $\tilde{\nu}$ = 2926 (w), 2856 (w), 1795 (w), 1741 (m), 1682 (w), 1605 (w), 1578 (w), 1481 (w), 1351 (s), 1319 (m), 1248 (s), 1197 (s), 1157 (m), 1103 (m), 958 (m), 897 (w), 825 (w), 794 (w), 749 (w), 717 (m), 683 cm^{-1} (w); HR-MALDI-MS (3-HPA): *m*/*z* (%): 1884.4403 (100, $[M+H]^+$, calcd for $C_{100}H_{89}I_2O_{16}N_6^+$: 1884.4458).



Cavitand 11

$[Pd(PPh_3)_4]$ (46 mg, 0.040 mmol) and CuI (8 mg, 0.04 mmol) were added to a degassed solution of **9** (161 mg, 0.398 mmol), **5** (0.75 g, 0.40 mmol), and DIPEA (1.4 mL, 8.2 mmol) in THF (30 mL). The mixture was stirred for 3 d at 25 °C, after which the solvent was evaporated. FC (SiO_2; $CH_2Cl_2 \rightarrow CH_2Cl_2$/EtOAc 97:3) and recycling GPC (Novogrom 100; CH_2Cl_2) afforded **11** (148 mg, 17 %) as a red solid. $R_f = 0.57$ (SiO_2; CH_2Cl_2/EtOAc 97:3); m.p. > 320 °C (decomp); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 0.84 - 0.93$ (m, 12H), 1.02 (t, $J = 7.5$, 6H), 1.19 – 1.47 (m, 38H), 2.02 – 2.15 (m, 11H), 6.12 – 7.13 (m, 4H), 2.18 (s, 6H), 2.24 (s, 3H), 2.34 (q, $J = 7.5$, 4H), 2.57 (s, 6H), 3.61 – 3.78 (m, 2H), 4.37 (t, $J = 7.4$, 2H), 7.32 – 7.36 (m, 2H), 7.37 – 7.54 (m, 6H), 7.57 (s, 1H), 7.60 (s, 1H), 7.66 – 7.71 (m, 2H), 7.71 – 7.77 (m, 4H), 8.07 – 8.22 ppm (m, 4H); ^{13}C NMR (101 MHz, $CDCl_3$): $\delta = 11.96$, 12.56, 13.97, 13.99, 14.63, 17.10, 17.74, 17.80, 18.08, 18.14, 22.52, 22.59, 26.96, 27.05, 29.05, 29.13, 31.50, 31.60, 31.67, 35.55, 37.70, 53.44, 89.67, 89.74, 96.21, 123.60, 124.65, 126.76, 128.57, 128.87, 129.20, 130.30, 130.56, 131.66, 131.78, 132.34, 132.96, 134.31, 136.07, 137.11, 137.31, 137.59, 137.67, 138.25, 138.83, 139.03, 139.21, 151.74, 154.05, 161.76, 161.92 ppm; ^{19}F NMR (377 MHz, $CDCl_3$): $\delta = -145.77$ ppm (q, $J = 33.0$, 2F); UV/Vis ($CHCl_3$): λ_{max} (ε) = 530 (59000); HR-MALDI-MS (3-HPA): m/z (%): 2139.7494 (100, $[M{-}F]^+$, calcd for $C_{125}H_{114}BFIN_8O_{16}^+$: 2139.7475).



Cavitand 2

[$Pd(PPh_3)_4$] (11 mg, 0.093 mmol) and CuI (1.7 mg, 0.0093 mmol) were added to a degassed solution of **12** (36 mg, 0.069 mmol), **11** (0.10 g, 0.046 mmol), and DIPEA (0.16 mL, 0.93 mmol) in THF (10 mL). The mixture was stirred for 3 d at 35 °C, after which the solvent was evaporated. FC (SiO_2; CH_2Cl_2 → CH_2Cl_2/EtOAc 97:3) and recycling GPC (Novogrom 100; CH_2Cl_2) afforded **2** (13 mg, 11 %) as a red solid. $R_f = 0.57$ (SiO_2; CH_2Cl_2/EtOAc 97:3); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 0.89$ (q, $J = 6.7$, 12H), 1.02 (t, $J = 7.5$, 6H), 1.19 – 1.43 (m, 38H), 1.45 (s, 6H), 2.11 (s, 8H), 2.18 (d, $J = 3.4$, 6H), 2.25 (d, $J = 3.4$, 6H), 2.35 (q, $J = 7.6$, 4H), 2.55 – 2.63 (m, 10H), 2.88 – 2.96 (m, 4H), 3.71 (t, $J = 7.2$, 2H), 4.38 (t, $J = 7.6$, 2H), 6.07 – 7.11 (m, 4H), 7.26 – 7.38 (m, 8H), 7.38 – 7.53 (m, 10H), 7.65 – 7.80 (m, 8H), 8.02 – 8.28 (m, 4H), 8.78 – 8.91 ppm (m, 2H); ^{13}C NMR (101 MHz, $CDCl_3$): $\delta = 11.96$, 12.41, 12.56, 13.98, 14.00, 14.63, 17.10, 18.09, 18.15, 20.49, 22.53, 22.59, 26.97, 27.06, 29.05, 29.14, 30.54, 31.51, 31.61, 35.57, 37.72, 53.44, 89.67, 89.74, 89.94, 123.60, 123.86, 124.63, 124.65, 126.77, 127.35, 128.08, 128.38, 128.58, 128.66, 129.04, 129.20, 129.25, 129.46, 130.31, 130.56, 131.68, 131.79, 132.17, 132.34, 132.43, 132.96, 133.75, 135.93, 136.08, 136.41, 137.11, 137.13, 137.33, 137.34, 138.25, 138.82, 140.70, 150.92, 151.75, 154.06, 161.93 ppm; ^{19}F NMR (377 MHz, $CDCl_3$): $\delta = -145.77$ (q, $J = 31.8$, 2F), −134.97 ppm (q, $J = 33.1$, 2F); IR (ATR): $\tilde{\nu} = 2927$ (w), 2857 (w), 1741 (m), 1680 (w), 1607 (w), 1579 (w), 1543 (w), 1480 (m), 1406 (m), 1336 (m), 1263 (m), 1188 (s), 1159 (m), 1084 (m), 959 (m), 897 (w), 796 (w), 774 (w), 715 (m), 648 cm^{-1} (w); UV/Vis ($CHCl_3$): $\lambda_{\max}$ (ε) = 530 (70000), 620 nm (114000); HR-MALDI-MS (3-HPA): *m*/*z* (%): 2537.0707 (100, $[M{-}F]^+$, calcd for $C_{160}H_{140}B_2F_3N_{10}O_{16}^+$: 2537.0621).

1.4 Quantification of Donor-Only Fractions in The Cavitand Samples

Cavitand **1a** had been initially published in 2005, and an unexpectedly high donor fluorescence intensity had been observed in its steady-state fluorescence spectrum.[4] The compound was resynthesized in 2010, together with compound **1b** and **1c**.[1] While the newly synthesized **1a** sample exhibited lower donor fluorescence intensity than the originally published one, it was still too high for assuming a close dye-dye distance.

However, the time-resolved fluorescence spectroscopy results of the current work supported the hypothesis of close dye-dye distances in all cavitands **1a**–**d**. Therefore, we suspicioned that donor-only fractions in the cavitand samples could explain the high donor fluorescence intensities in the steady state fluorescence spectra.

We assumed that the reason for the presence of donor-only fractions is the loss of the BF_2 unit of the acceptor dye during the course of cavitand synthesis, presumably during a TMS-promoted deprotection step of a cavitand dye arm precursor. The larger instability of the acceptor dye compared to the donor dye had been noticed in the SI of reference [2].[1] Loss of the BF_2 unit in BODIPY dyes has precedence and was observed in high acidity media,[5] or under strongly basic conditions.[6] Hints for the partial lack of the BF_2 unit were found in MALDI-MS spectra in the form of $[M\text{–}BF_2]^+$ signals (which, however, might also form during the ionization in the mass spectrometer), while no evidence could be drawn from ^{1}H, ^{13}C, and ^{19}F NMR spectra published in reference [2] for cavitands **1a**–**c**.[1]

Therefore, we obtained new ^{19}F NMR spectra using special conditions (pulse sequence with a 30° flip angle) that allowed quantitative integration of the BF_2 signals, corresponding to the donor and acceptor moieties (see Figure S2). These spectra revealed that cavitands **1a** and **1b** indeed posses significant donor-only portions (18% and 11%, respectively), while cavitands **1c**, **1d**, and **2** show very small donor-only portions that are smaller than the NMR integration error of up to 2%. Due to the small molecular mass difference between the fractions equipped with and lacking the BF_2 unit, separation with gel permeation chromatography (GPC) was not possible. Neither was silica gel chromatography possible due to negligible polarity differences.

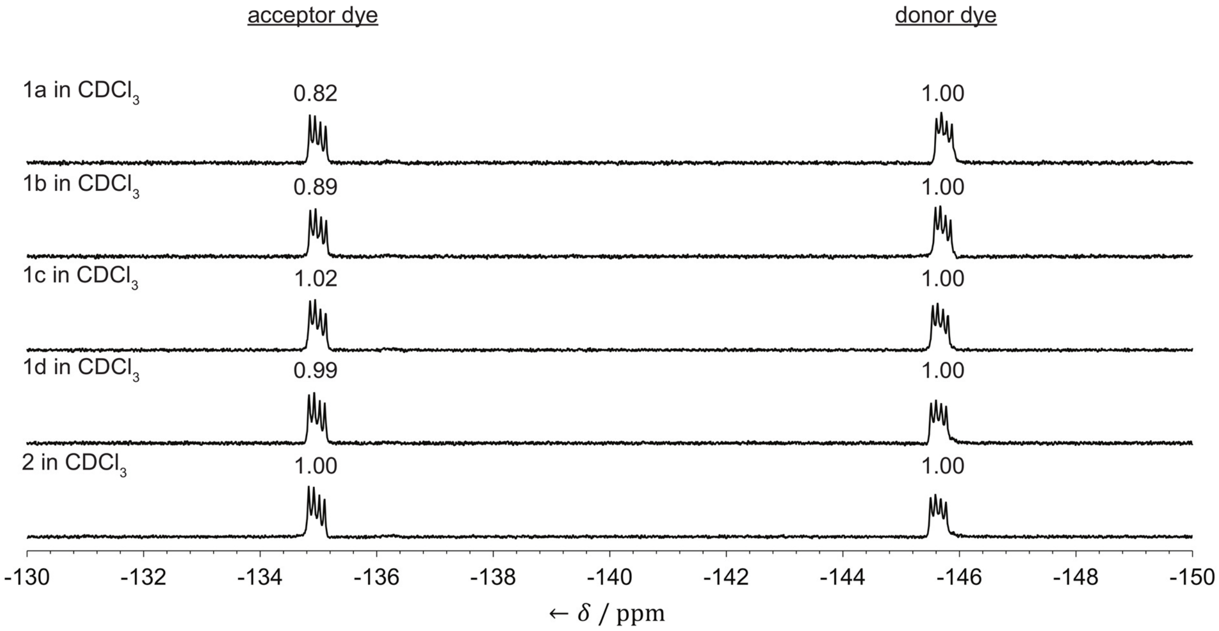


Figure S2. ^{19}F NMR spectra (298 K, 376 MHz) of cavitands **1a**–**d** and **2** in $CDCl_3$ employing a 30° flip angle.

2 X-Ray Data of Cavitand 5

Crystal data for cavitand **5** were deposited with the Cambridge Crystallographic Data Base with CCDC number 953763 ($CDCl_3$, kite), and can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

A clear yellow prism-like specimen of $C_{115}H_{103}Cl_{45}I_2N_6O_{16}$ (formula weight: 3674.08), approximate dimensions 0.140 mm × 0.240 mm × 0.300 mm, was used for the X-ray crystallographic analysis at 100(2) K. The X-ray intensity data were measured on a Bruker Kappa Apex-II Duo system equipped with a graphite monochromator (λ = 0.71073 Å).

The integration of the data using a monoclinic unit cell yielded a total of 239167 reflections to a maximum θ angle of 27.55° (0.77 Å resolution), of which 67783 were independent (average redundancy 3.528, completeness = 98.9%, R_{int} = 7.89%, R_{sig} = 11.35%) and 33738 (49.77%) were greater than $2\sigma(F^2)$. The index range was: $-35 \leq h \leq 36$, $-37 \leq k \leq 27$, $-49 \leq l \leq 49$. The final cell constants of a = 27.741(2) Å, b = 29.059(3) Å, c = 38.382(3) Å, β = 106.309(4)°, V = 29696.(4) Å^3, are based upon the refinement of the XYZ-centroids of 9104 reflections above 20 $\sigma(I^2)$ with $4.422° < 2\theta < 53.93°$. Data were corrected for absorption effects using the multi-scan method (SADABS). The ratio of minimum to maximum apparent transmission was 0.796. The absorption coefficient is 1.293 mm^{-1}.

The structure was solved by direct methods and refined using the OLEX2 and Bruker SHELXTL Software Package, using the space group P 1 21/n 1, with Z = 8 for the formula unit $C_{115}H_{103}Cl_{45}I_2N_6O_{16}$. The final anisotropic full-matrix least-squares refinement on F^2 with 2049 variables converged at $R1$ = 22.48% ($wR2$ = 54.52%), for the observed data and $R1$ = 32.80% ($wR2$ = 58.43%) for all data. The goodness-of-fit was 3.173. The largest peak in the final difference electron density synthesis was 8.089 e Å^{-3} and the largest hole was −5.517 e Å^{-3} with an RMS deviation of 0.419 e Å^{-3}. On the basis of the final model, the calculated density was 1.644 g cm^{-3} and F(000), 14672 e.

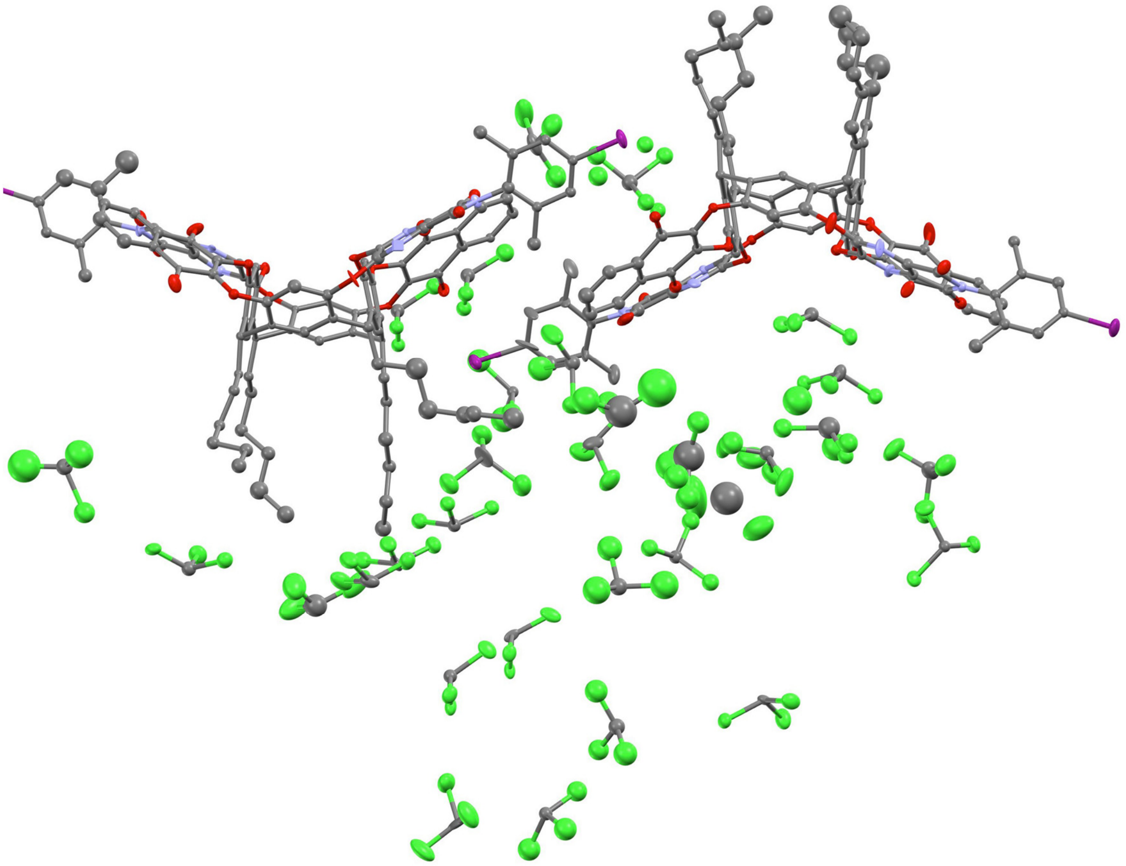


Figure S3. Asymmetric unit of the crystal structure of cavitand **5** (kite, from $CDCl_3$) measured at 100 K. Thermal ellipsoids are shown at the 50% probability level.

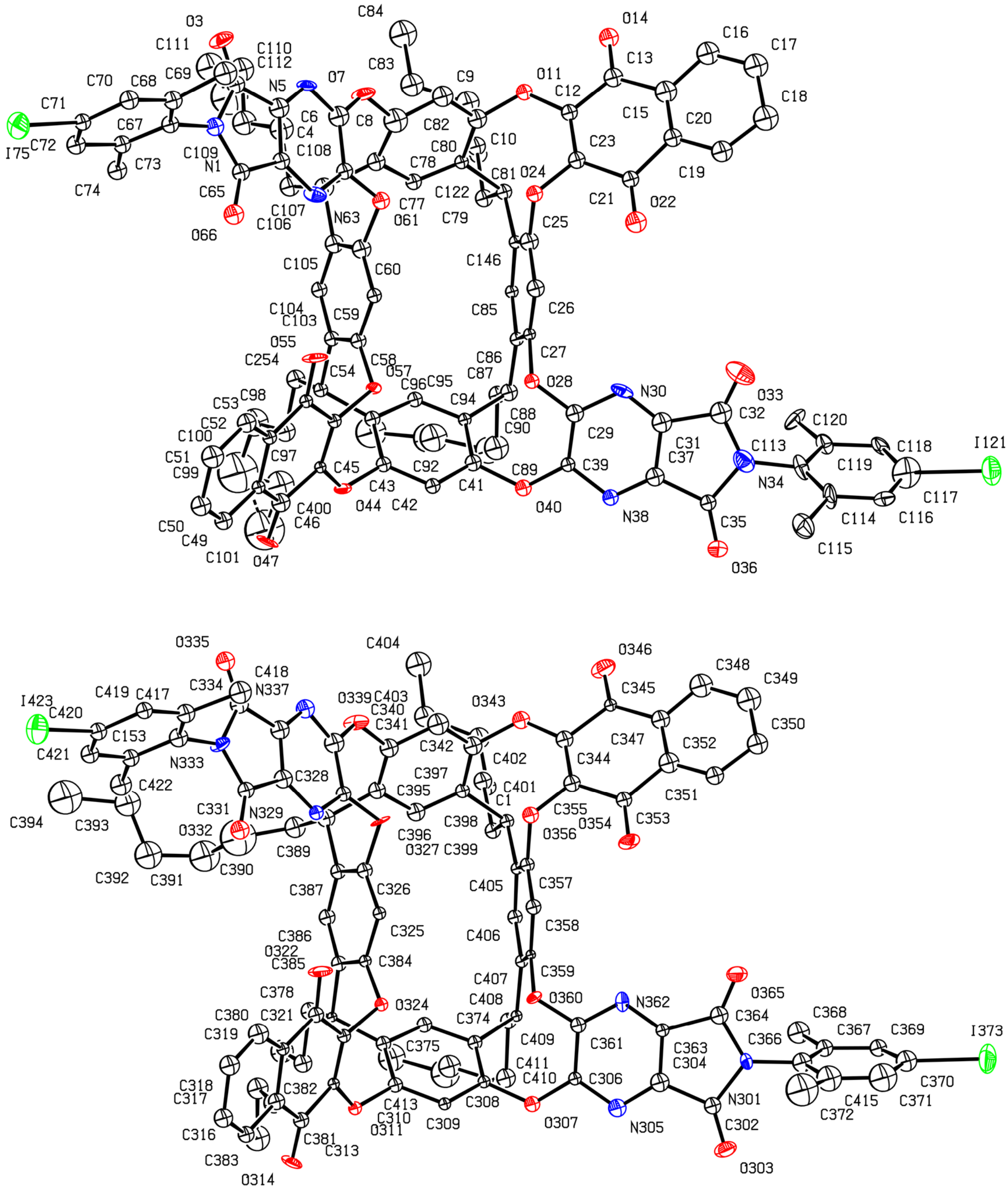


Figure S4. Molecular structures of **5** in the crystal including atom numbers (kite, from $CDCl_3$) measured at 100 K. Thermal ellipsoids are shown at the 50% probability level.

3 Fluorescence Section

3.1 Reference Donor and Acceptor BODIPY Dye Fluorescence Lifetimes

Fluorescence lifetime decays were measured at the University of Zurich using a custom-built instrument described previously.[7] Picosecond light pulses from a white light source (SC-450-4, 20 MHz, Fianium, Southampton, UK) were used for excitation. The excitation wavelengths were selected by HQ470/40 (Chroma) and z582/15 (Semrock) bandpass filters. The binning width of the recorded histograms is 4 ps.

Cavitands **3** and **1a** were employed to measure the mean fluorescence lifetimes τ_D and τ_A. The donor dye decay curve of cavitand **3** after donor excitation with $\lambda_\mathrm{exc} = 470$ nm, and the acceptor dye decay curve of cavitand **1a** after acceptor excitation with $\lambda_\mathrm{exc} = 582$ nm are presented below. The curves were fitted with monoexponential decay functions convolved with the instrument response functions (IRF), yielding the mean fluorescence lifetimes $\tau_\mathrm{D} = 4.21$ ns and $\tau_\mathrm{A} = 5.24$ ns.

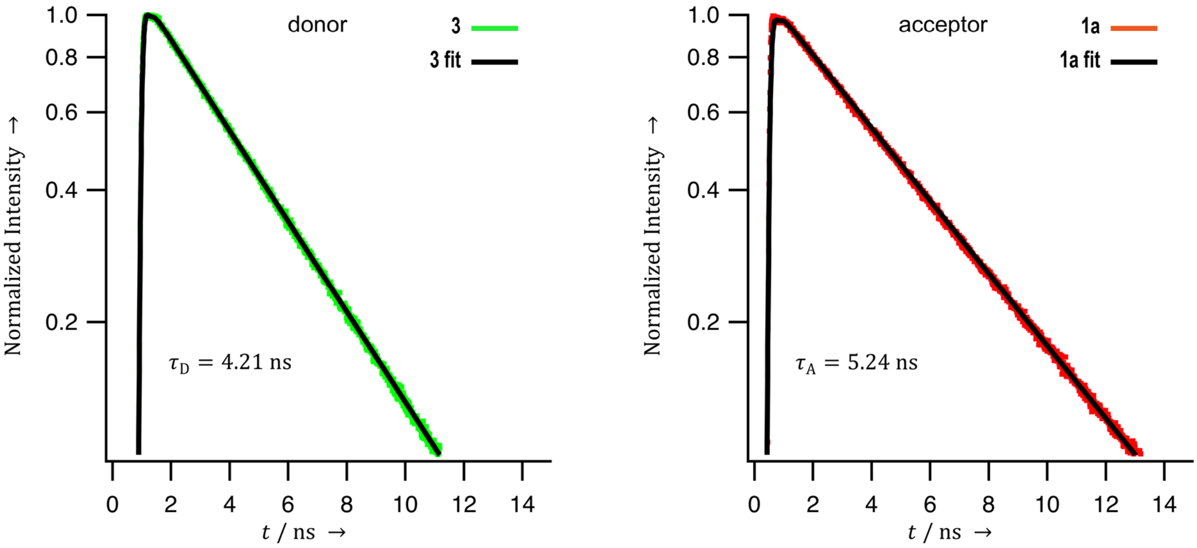


Figure S5. Fluorescence decay curves of reference donor cavitand **3** (left) excited with $\lambda_\mathrm{exc} = 470$ nm, and selectively excited acceptor of cavitand **1a** (right) with $\lambda_\mathrm{exc} = 582$ nm. Measurements were performed under magic angle configuration i.e. emission polarizer set to 54.7° with respect to excitation polarization. Monoexponential decay fitting yielded mean fluorescence lifetimes $\tau_\mathrm{D} = 4.21$ ns for the donor BODIPY dye and $\tau_\mathrm{A} = 5.24$ ns for the acceptor BODIPY dye.

3.2 Discussion on Potential Electron Transfer

While electron-poor quinones can act as electron acceptors, in case of our quinone moieties the acceptor property is diminished due to electron-donation from the O-atoms of the cavitand backbone; the first redox potential of the quinone moiety was recorded at −1.1 V.[8] In addition, owing to the stiffness of the linker, direct contact between the quinone walls and the dyes through space is essentially impossible, and electron transfer through the linker arm is very unlikely. Electron transfer from the quinone moieties to the excited dyes would also be accompanied by broad, red-shifted charge-transfer bands in the fluorescence spectra of the cavitands. The absence of such bands is another argument against the occurrence of electron transfer in these molecules. Finally, Figure S6 below shows quantum yields (QY) of relevant single BODIPY dyes **6**–**8** and cavitands **1a**–**d** (Details on the determination of the quantum yields are reported in Section 3.4 of the Supporting Information). The QYs are independent of the length of the linker, which also strongly argues against electron transfer to the quinone walls. In summary, electron transfer from the chromophoric group to the quinone groups does not have to be taken into account for our analysis.

Concerning potential photoinduced electron transfer between the two BODIPY dyes, this energetics can be estimated with the Rehm-Weller-equation, which takes into account the redox potentials of the donor and acceptor dyes. The redox potentials of the BODIPY dyes used in this study were previously measured (see Figure S6 below). Further required parameters are the excited state energy of the $S_0 \rightarrow S_1$ transition of the donor (E_{00} = 2.33 eV, calculated from the wavelength of the absorption maximum of the donor dye, λ_{max} = 529 nm) and a work term of 0.3 eV (empirical coulombic factor for non-polar solvents). Using these data, we obtain a Gibbs free energy of ΔG = 1.2 kcal mol^{-1} for the photoinduced electron transfer from the excited donor to the acceptor dye. On the other hand, we obtain ΔG = 3.9 kcal mol^{-1} for the photoinduced electron transfer from the acceptor to the excited donor. Both processes are thus endergonic. Not only is electron transfer between the donor and acceptor BODIPY dyes in the closed cavitands unlikely due to thermodynamic reasons, electron transfer has even more stringent requirements for the distance between the two dyes (i.e. if electron transfer does take place, Förster transfer will be extremely efficient also). Possible electron transfer would therefore not have any bearing on the central conclusion regarding the vase cavitands, which is that the opening angle is close to zero. In the kite cavitand, the distance between donor and acceptor is clearly too large for electron transfer to occur.

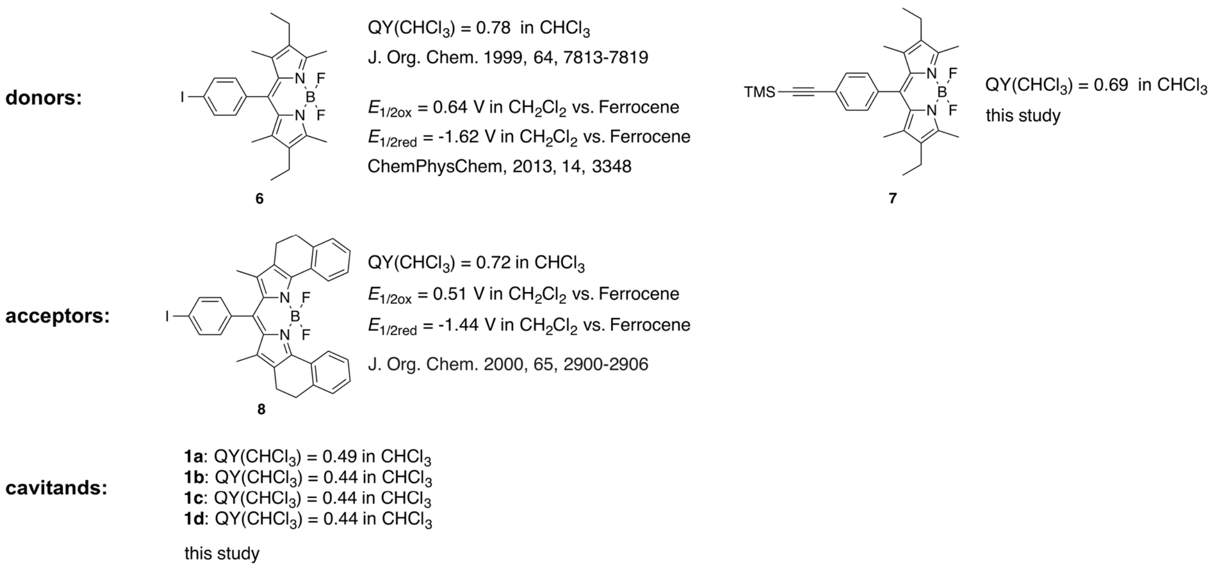


Figure S6. Quantum yields and redox potentials of relevant BODIPY dyes **6**–**8** and cavitands **1a**–**d**.

3.3 Determination of Quantum Yields and Förster Radius R_0

Measurements of quantum yields (QY) and determination of R_0 were carried out in the Beckman Institute Laser Resource Center (California Institute of Technology) and were supported by the Arnold and Mabel Beckman Foundation. All QY's were measured at concentrations of <1 μM in chloroform at room temperature and in ambient air. Solid samples were stored in the dark and refrigerated, solutions were prepared directly before taking data. Maximum peak absorptions did not exceed 0.25 to avoid detrimental self-absorption effects. UV/Vis absorption spectra were collected with a Cary 50 UV/Vis spectrophotometer in 1 cm pathlength quartz cuvettes. The acceptor extinction coefficient was determined in the following way: Four small amounts of acceptor (**9**, Figure S8) were weighed out and dissolved in known volumes of chloroform. The optical spectra of these solutions of known concentrations were measured and the average extinction coefficient of **9** was calculated using Beer's law. Emission spectra were recorded on a Jobin Yvon Spec Fluorolog-3-11. Samples were excited with 532-nm light, which was provided by a xenon arc lamp equipped with a monochromator for wavelength selection. Right angle emission was diffracted with a monochromator and detected with a Hamamatsu R928P photomultiplier tube with photon counting.

QY data of samples **1a**–**d** and **7** were collected using the comparative method of Williams *et al.*[9] This method involves the use of a well-characterized standard sample (here: anthracene in ethanol) with a known QY value. The measured QY data are summarized in Figure S7.

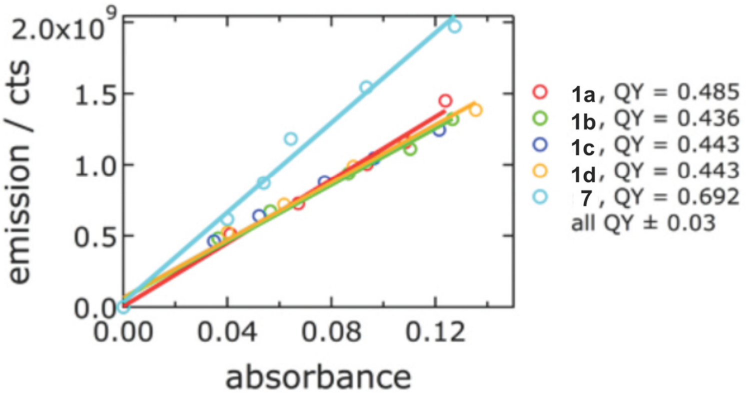


Figure S7. Measured integrated fluorescence intensity vs. absorbance (circles). The solid lines are linear fits with the slope *m* and the intercept 0.

QY's were derived by comparison to a known value in the following way: The slopes *m* of the unknown substances (*x*) and standard (*std*) were measured as described above. The QY of an unknown substance (*x*) is given by the equation below, in which *n* denotes the index of refraction of the solvent.

$$QY_x = QY_{std} \cdot \left(\frac{m_x}{m_{std}}\right) \cdot \left(\frac{n_x^2}{n_{std}^2}\right)$$

The Förster radius R_0 was determined using the equations below and the measured acceptor extinction spectrum (experimental details see above).

Donor: $R_0^6 = \dfrac{9\,\mathrm{QY}_D(\ln 10)\kappa^2 J}{128\,\pi^5 n^4 N_A}$

$\kappa^2 = 2/3$ random orientation of donor molecules in solution

$n = 1.4460$ refractive index of chloroform

N_A : Avogadro's number

$$J = \int f_D(\lambda)\epsilon_A(\lambda)\lambda^4 \,\mathrm{d}\lambda$$

f_D : normalized donor fluorescence spectrum

ϵ_A : acceptor molar extinction spectrum

λ : wavelength

The acceptor (**9**) extinction and normalized donor (**7**) emission spectra are depicted in Figure S8.

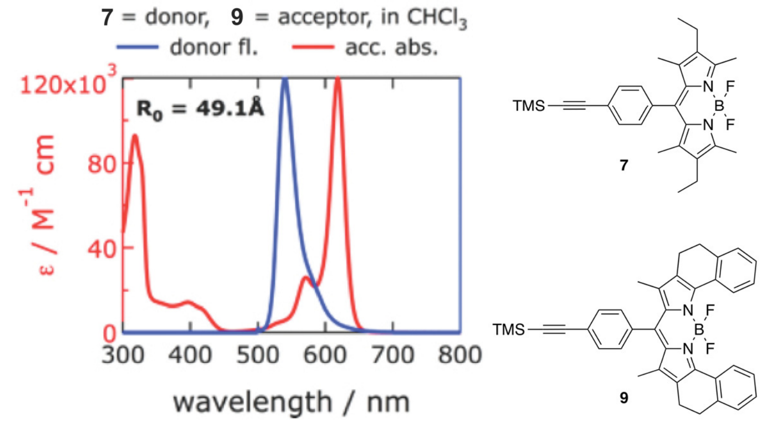


Figure S8. Acceptor extinction (**9**, red) and normalized donor emission (**7**, blue) spectra.

A Förster radius R_0 = 49.1 Å was obtained from steady-state absorption and emission spectra, and the equation above was used; κ^2 was assumed to be 2/3, which is a good approximation for a random orientation of donor and acceptor molecules in solution.

3.4 Analysis of Decay Curves

The fluorescence decay curves of the donor dyes in cavitands **1a**–**d** show multiexponential behavior whose description requires at least two decay times. The results of biexponential fits are shown in the Table below.

Table S1. Biexponential fits of the fluorescence decay curves of the donor moieties of cavitands **1a**–**d**

cavitand	τ_1 / ns	τ_2 / ns
1a	0.3	3.8
1b	0.5	3.9
1c	0.5	3.7
1d	0.8	3.8

Fitted with IRF-convolution

4 MD Setup

4.1 Preparation of GAFF Input Files

Cavitands and linker molecules were assembled from HF/6-31G(d)-optimized fragments. Atom charges were calculated with the restrained electrostatic potential (RESP) fitting method of the molecular electrostatic potential (MEP),[10] performed on the RED Server[11] (development version 2012; allows to input up to 350 atoms) with the charge model RESP-A1A (HF/6-31G(d)//HF/6-31G(d) – Connolly surface algorithm used in MEP computation – 2 stage RESP fit qwt = 0.0005/0.001), using the Gaussian 09 software[12]. Bonded and non-bonded parameters were assigned with the program ACPYPE,[13] which interfaces with the program ANTECHAMBER.[14] Bonded and non-bonded parameters involving the boron atoms were taken from parameters derived for CHARMM,[15] and ported to GROMACS by means of unit conversion. Chloroform parameters (GAFF) and liquid structure coordinates were obtained from virtualchemistry.org.[16]

4.2 Preparation of CGenFF Input Files

Cavitands and linker molecules were parameterized for CGenFF (version 2b7) via ParamChem Web service (www.paramchem.org, version 0.9.6).[17] Bonded and non-bonded parameters involving the boron atoms were taken from literature.[15] Chloroform parameters (OPLS) and liquid structure coordinates were obtained from virtualchemistry.org.[16]

4.3 MD Run Parameters.

Cavitands or linker molecules, respectively, were solvated in a box of pre-equilibrated chloroform molecules with a minimal distance of 13 Å between any atom of the solute and the periodic boundary. As the guest-exchange rate for 2,6-dimethylphenyl-substituted cavitands is slow on the NMR timescale,[18] we expected that chloroform molecules would not diffuse into the cavity on the MD simulation timescale by itself. Therefore, one chloroform molecule was manually placed into the cavities of cavitands **1a**–**d**. It was observed that this chloroform molecule remained in the cavity throughout the whole simulation. The simulation time was 500 ns for cavitands and 100 ns for linker molecules, with a time step of 1 fs (such a short time step is required due to the high vibrational frequency of triple bonds present in linker units). Cavitand coordinates were saved every 4 ps and linker molecule coordinates every 5 ps. Long-range electrostatics and van der Waals interactions were treated with a simple cut-off scheme using PME and a cut-off at 10 Å.[19] The temperature was kept constant at 298 K by applying the Velocity Rescaling algorithm.[19] The system pressure was kept constant at 1 atm with the Parinello Rahman Barostat.[19] The LINCS algorithm was used to keep all bonds involving hydrogen atoms constrained.[19] All molecular dynamics simulations were performed with a Message Passing Interface (MPI) version of GROMACS v. 4.5.5.[20] Simulations were performed on the Brutus super computer,[21] employing 16 cores per simulation. The total simulation time accumulated over all simulations was 5 µs for the cavitand and 0.8 µs for the linker molecules. Atom coordinates were extracted from each snapshot via the g_traj routine of GROMACS, and subsequent calculations were performed with the Python programming language version 2.7.

5 Fluorescence Decay Curves from MD Data

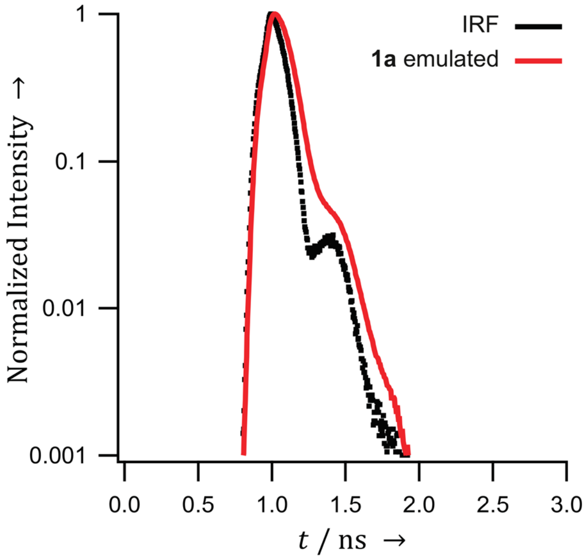
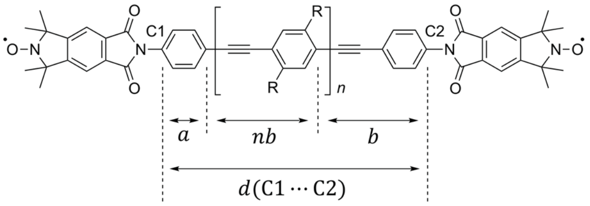


Figure S9. Donor fluorescence emission decay curve of vase cavitand **1a** with an implemented donor-only fraction of 0% emulated on the basis of a CGenFF-simulation trajectory: an extremely rapid decay is observed that basically parallels the IRF, if a donor-only fraction is not taken into account for the emulation.

6 Linker Study

We subjected molecules **13a**–**d**[22] (Figure S10, $O^{\cdot}$ was replaced by H) to MD simulations with CGenFF and GAFF (explicit chloroform, 100 ns each).



13a: *n* = 1
13b: *n* = 2
13c: *n* = 3
13d: *n* = 4

Figure S10. Test compounds **13a–d** used in literature to determine the persistence length of a phenylene-ethynylene-based polymer to be $L_\mathrm{p} = 13.8 \pm 1.5$ nm.[22] We used these test compounds to determine and compare the persistence lengths resulting from MD simulations with CGenFF and GAFF.

The obtained $d(\mathrm{C1} \cdots \mathrm{C2})$ distance distribution histograms are illustrated in Figure S11. The histograms obtained with CGenFF are broader than the ones obtained with GAFF, indicating that GAFF parameters result in stiffer phenylene-ethynylene linkers than CGenFF parameters. Quantification of L_p values from MD data can be achieved by applying the Worm-Like Chain (WLC) model, which is often used to describe the behavior of semi-flexible polymers.[23] According to the WLC model, the mean square end-to-end distance $\langle R^2 \rangle = \langle d(\mathrm{C1} \cdots \mathrm{C2}) \rangle$ of a chain can be described by Equation:[22a]

$$\langle R^2 \rangle = 2 c_\mathrm{l} L_\mathrm{p} \left(1 - \frac{L_\mathrm{p}}{c_\mathrm{l}} \left(1 - e^{-\frac{c_\mathrm{l}}{L_\mathrm{p}}} \right) \right) \qquad (1)$$

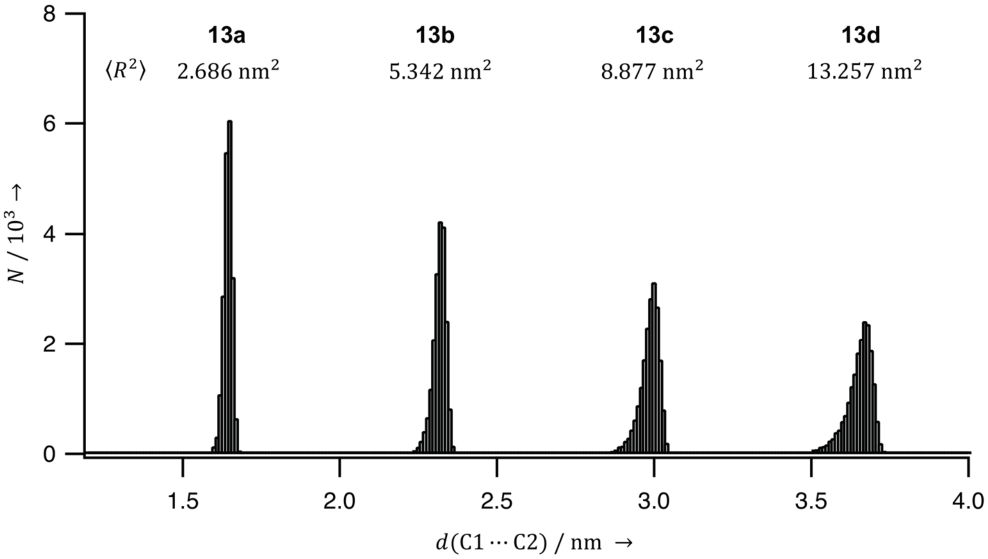
where c_l is the contour length – the $\langle d(\mathrm{C1} \cdots \mathrm{C2}) \rangle$ distance of the chain in its stretched, linear form. The contour lengths of systems **13a**–**d** are composed of one phenyl unit with the length a, and $(n+1)$ phenylene-ethynylene units with the length b (Figure S10). Thus, the compound-specific c_l values of **13a**–**d** can be described according to:

$$c_l(n) = a + (n+1)b \qquad (2)$$

The parameters a and b are force field-specific and were therefore used as fit parameters together with L_p. Substituting Equation (2) in (1) yields:

$$\langle R^2\rangle(n) = 2(a+(n+1)b)L_p\left(1-\frac{L_p}{a+(n+1)b}\left(1-e^{-\frac{a+(n+1)b}{L_p}}\right)\right) \qquad (3)$$

Equation (3) was used to fit the $\langle R^2\rangle(n)$ values obtained by CGenFF and GAFF, with a, b, and L_p as fit parameters. The resulting plots and fits are illustrated in Figure S12.



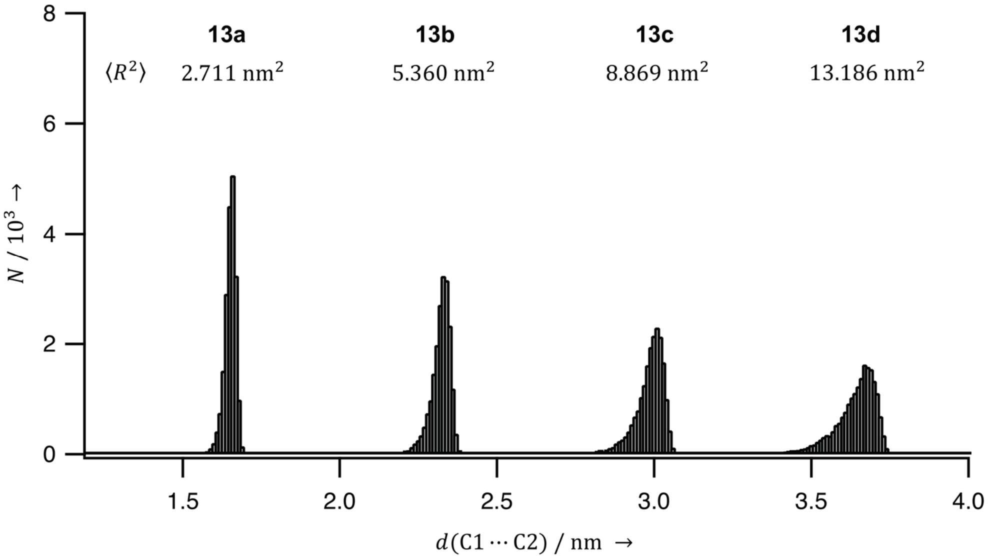


Figure S11. Histograms of distances $\langle d(\mathrm{C1}\cdots\mathrm{C2})\rangle$ in test compounds **13a**–**d** obtained from MD simulations with GAFF (top) and CGenFF (bottom). $\langle R^2\rangle = \langle d(\mathrm{C1}\cdots\mathrm{C2})\rangle$.

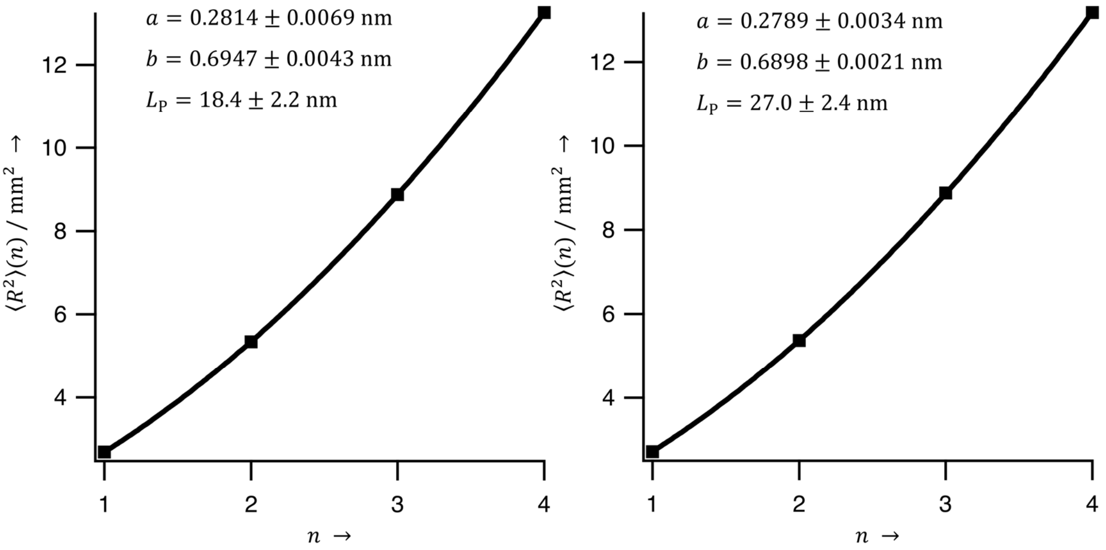


Figure S12. Plots of the mean square end-to-end distances $\langle R^2 \rangle$ against the number of phenylene-ethynylene linker units n in test compounds **13a**–**d** simulated with CGenFF (left) and GAFF (right). The plots were fitted to Equation (3) to determine the contour lengths L_p.

7 NMR Spectra of the Products

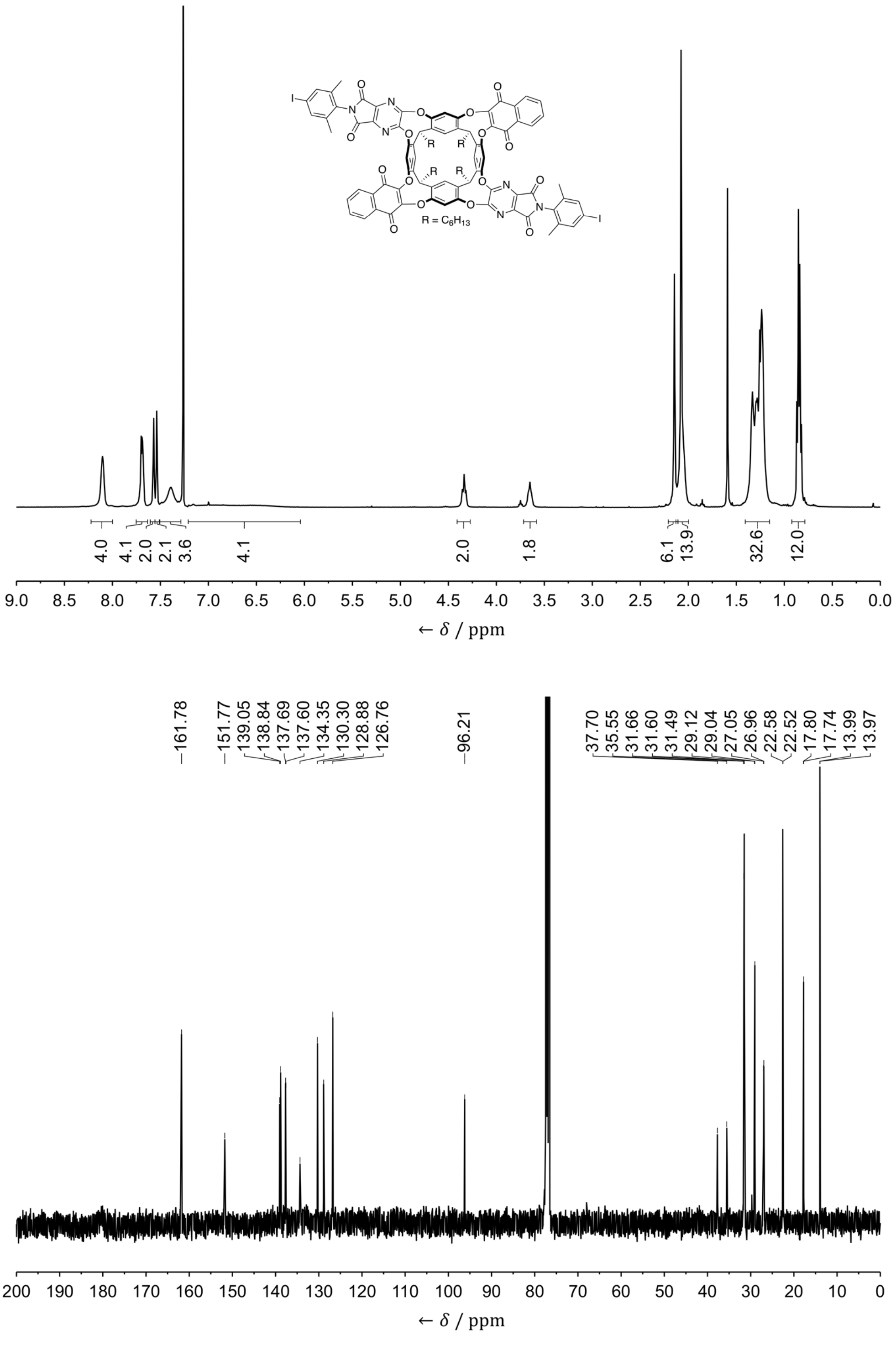


Figure S13. ^{1}H NMR (top, 400 MHz) and ^{13}C NMR (bottom, 100 MHz) spectra of compound **5** in CDCl$_3$ at 298 K.

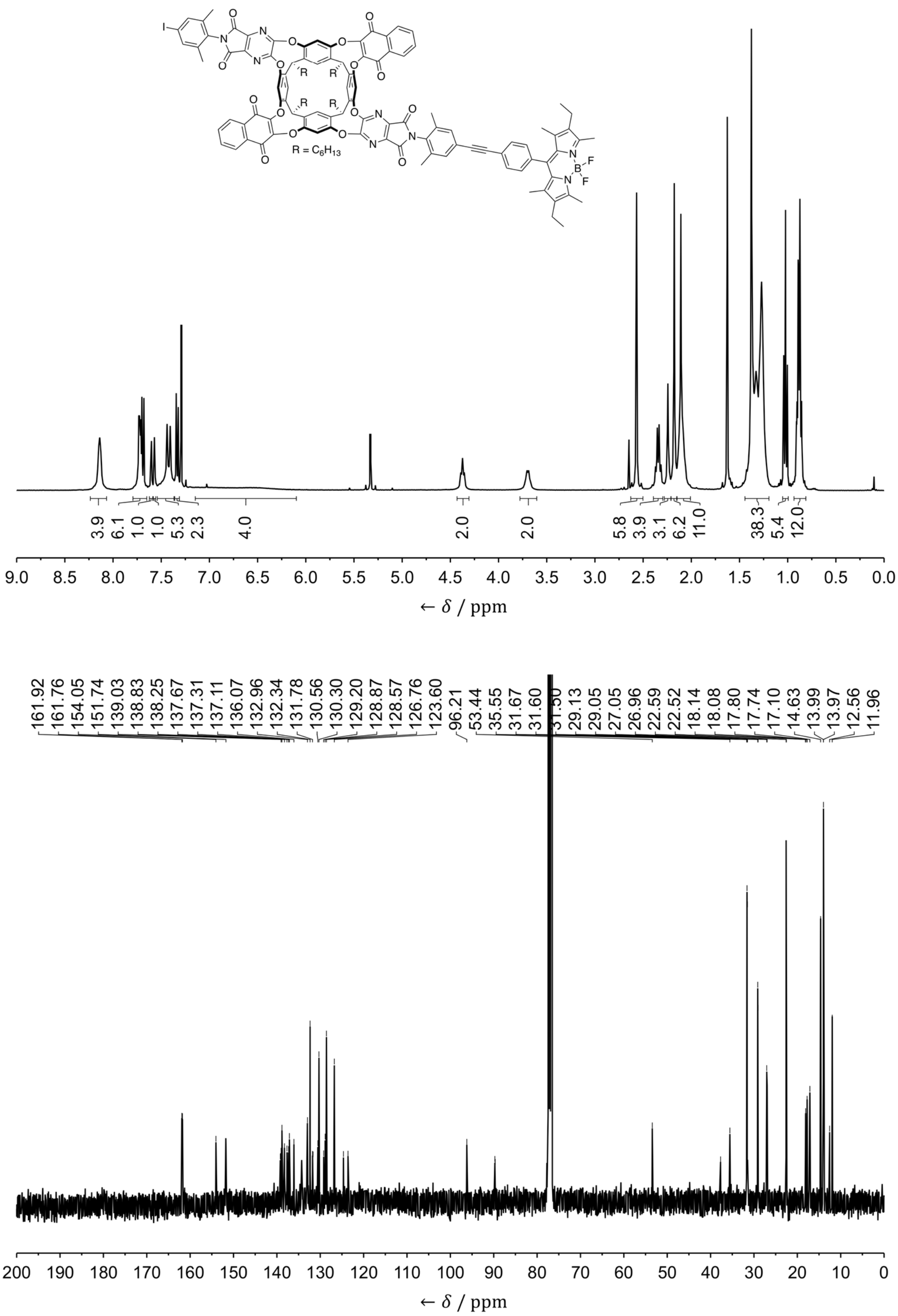


Figure S14. ^{1}H NMR (top, 400 MHz) and ^{13}C NMR (bottom, 100 MHz) spectra of compound **11** in CDCl$_3$ at 298 K.

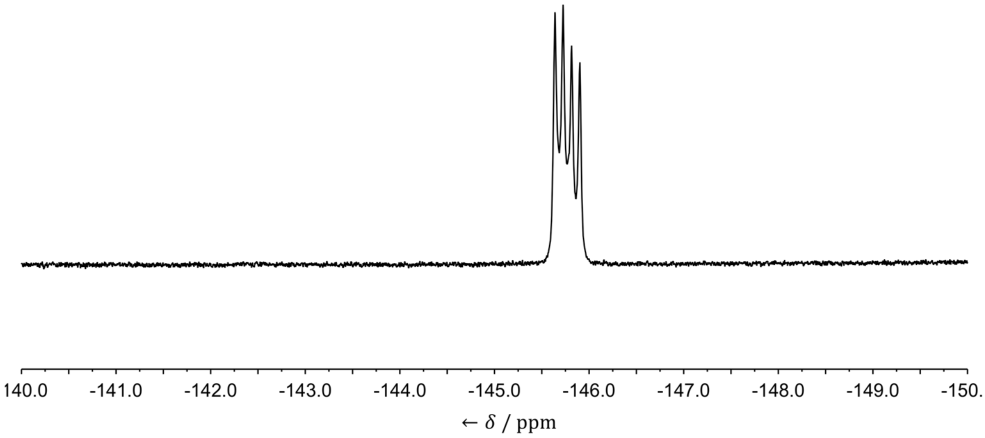


Figure S15. ^{19}F NMR (376 MHz) spectrum of compound **11** in $CDCl_3$ at 298 K.

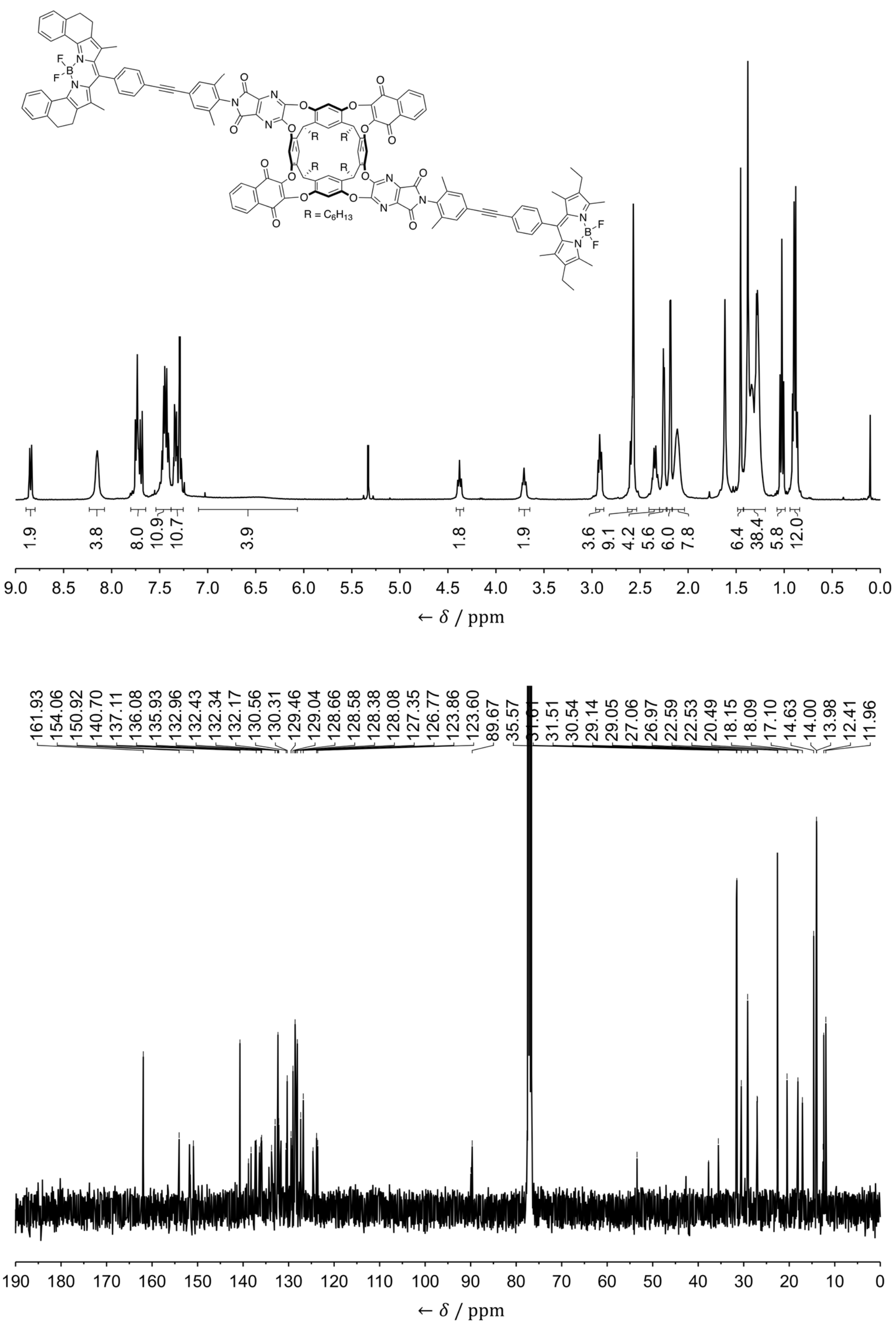


Figure S16. ^{1}H NMR (top, 400 MHz) and ^{13}C NMR (bottom, 100 MHz) spectra of compound **2** in CDCl$_3$ at 298 K.

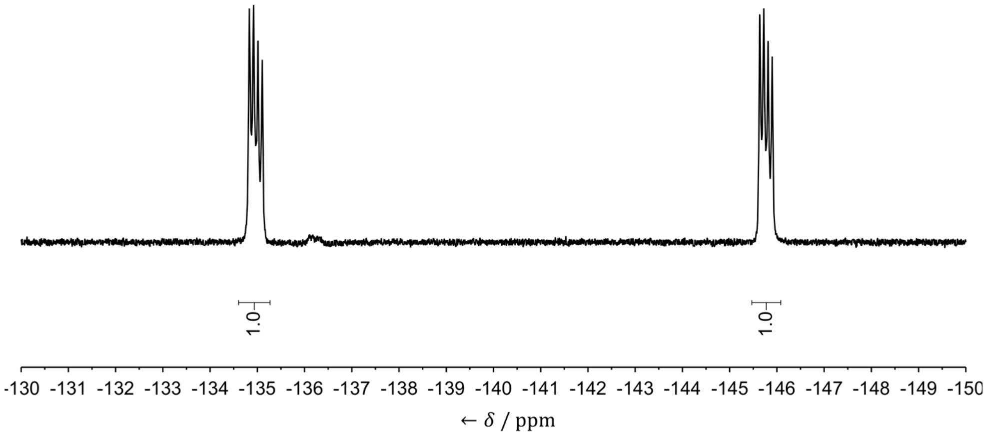


Figure S17. ^{19}F NMR (376 MHz) spectrum of compound **2** in $CDCl_3$ at 298 K.

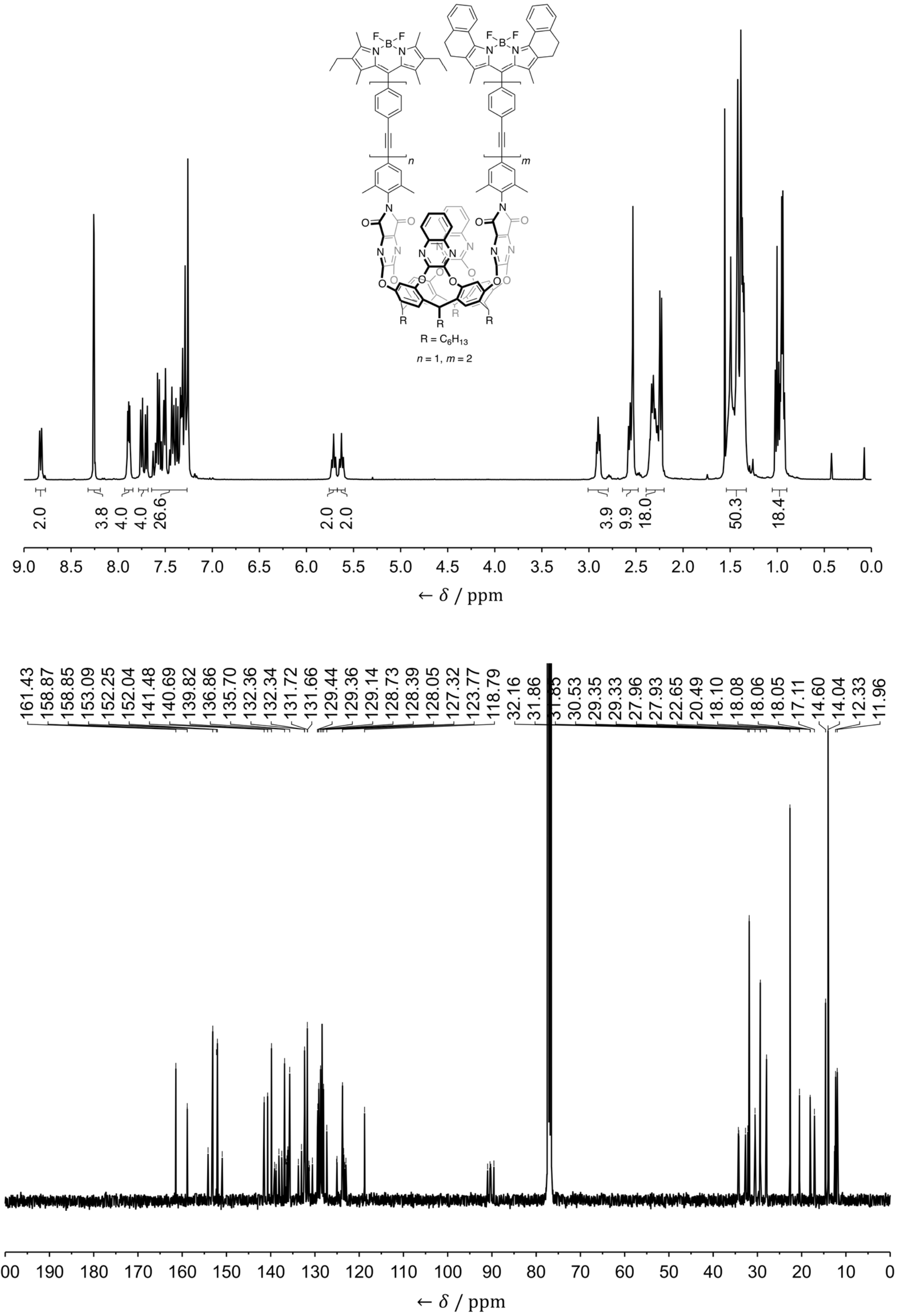


Figure S18. ^{1}H NMR (top, 400 MHz) and ^{13}C NMR (bottom, 100 MHz) spectra of compound **1d** in CDCl$_3$ at 298 K.

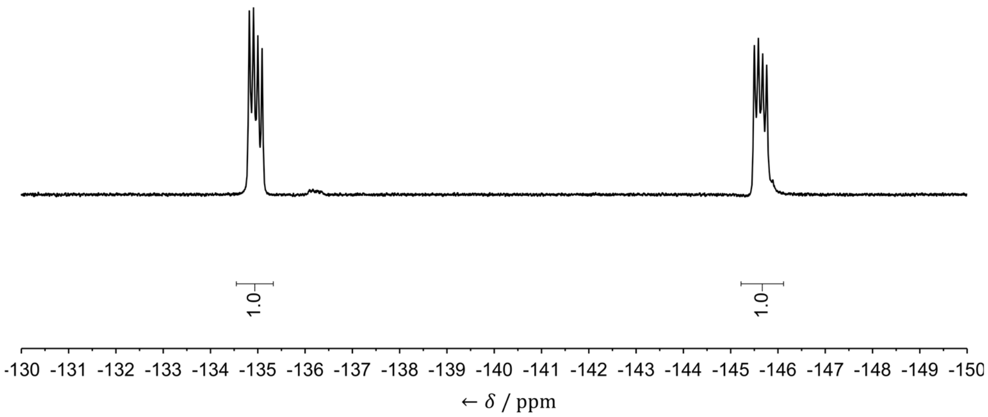


Figure S19. ^{19}F NMR (376 MHz) spectrum of compound **1d** in $CDCl_3$ at 298 K.

8 References

(1) Pochorovski, I.; Breiten, B.; Schweizer, W. B.; Diederich, F., *Chem. Eur. J.* **2010**, *16*, 12590.

(2) Azov, V. A.; Schlegel, A.; Diederich, F., *Bull. Chem. Soc. Jpn.* **2006**, *79*, 1926.

(3) Pochorovski, I.; Ebert, M.-O.; Gisselbrecht, J.-P.; Boudon, C.; Schweizer, W. B.; Diederich, F., *J. Am. Chem. Soc.* **2012**, *134*, 14702.

(4) Azov, V. A.; Schlegel, A.; Diederich, F., *Angew. Chem. Int. Ed.* **2005**, *44*, 4635.

(5) (a) Liras, M.; Prieto, J. B.; Pintado-Sierra, M.; Arbeloa, F. L.; García-Moreno, I.; Costela, Á.; Infantes, L.; Sastre, R.; Amat-Guerri, F., *Org. Lett.* **2007**, *9*, 4183; (b) Bañuelos, J.; López Arbeloa, F.; Arbeloa, T.; Salleres, S.; Vilas, J. L.; Amat-Guerri, F.; Liras, M.; López Arbeloa, I., *J. Fluoresc.* **2008**, *18*, 899.

(6) (a) Crawford, S. M.; Thompson, A., *Org. Lett.* **2010**, *12*, 1424; (b) Smithen, D. A.; Baker, A. E. G.; Offman, M.; Crawford, S. M.; Cameron, T. S.; Thompson, A., *J. Org. Chem.* **2012**, *77*, 3439.

(7) Nettels, D.; Hoffmann, A.; Schuler, B., *J. Phys. Chem. B* **2008**, *112*, 6137.

(8) Pochorovski, I.; Boudon, C.; Gisselbrecht, J.-P.; Ebert, M.-O.; Schweizer, W. B.; Diederich, F., *Angew. Chem. Int. Ed.* **2012**, *51*, 262.

(9) Williams, A. T. R.; Winfield, S. A.; Miller, J. N., *The Analyst* **1983**, *108*, 1067.

(10) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., *J. Phys. Chem.* **1993**, *97*, 10269.

(11) (a) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821; (b) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y., *Nucl. Acids Res.* **2011**, *39*, W511.

(12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.;

Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09 Revision A.1, Gaussian, Inc., Wallingford CT, 2009.

(13) Sousa da Silva, A.; Vranken, W., *BMC Res. Notes* **2012**, *5*, 367.

(14) (a) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *J. Comput. Chem.* **2004**, *25*, 1157; (b) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., *J. Mol. Graph. Model.* **2006**, *25*, 247.

(15) Song, K. C.; Livanec, P. W.; Klauda, J. B.; Kuczera, K.; Dunn, R. C.; Im, W., *J. Phys. Chem. B* **2011**, *115*, 6157.

(16) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D., *J. Chem. Theory Comput.* **2012**, *8*, 61.

(17) (a) Vanommeslaeghe, K.; Mackerell, J., A D, *J. Chem. Inf. Model.* **2012**, *52*, 3144; (b) Vanommeslaeghe, K.; Raman, E. P.; Mackerell, J., A D, *J. Chem. Inf. Model.* **2012**, *52*, 3155.

(18) Shirtcliff, L. D.; Xu, H.; Diederich, F., *Eur. J. Org. Chem.* **2010**, 846.

(19) van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; E. Apol, P. J. M.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C., Gromacs User Manual version 4.5.5.

(20) (a) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theory Comput.* **2008**, *4*, 435; (b) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., *J. Comput. Chem.* **2005**, *26*, 1701; (c) Lindahl, E.; Hess, B.; van der Spoel, D., *J. Mol. Model.* **2001**, *7*, 306; (d) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Comput. Phys. Commun.* **1995**, *91*, 43.

(21) Brutus Cluster. https://www1.ethz.ch/id/services/list/comp_zentral/cluster/index_EN.

(22) (a) Jeschke, G.; Sajid, M.; Schulte, M.; Ramezanian, N.; Volkov, A.; Zimmermann, H.; Godt, A., *J. Am. Chem. Soc.* **2010**, *132*, 10107; (b) Godt, A.; Schulte, M.; Zimmermann, H.; Jeschke, G., *Angew. Chem. Int. Ed.* **2006**, *45*, 7560.

(23) Kratky, O.; Porod, G., *Recl. Trav. Chim. Pay-B* **1949**, *68*, 1106.

Chapter 3

Concurrent Cooperativity and Substrate Inhibition in the Epoxidation of carbamazepine by Cytochrome P450 3A4 Active Site Mutants Inspired by Molecular Dynamics Simulations

Christian S. Müller, Tim Knehans, Dmitri R. Davydov, Patricia L. Bounds, Ursula von Mandach, James R. Halpert, Amedeo Caflisch, and Willem H. Koppenol

Biochemistry 2015, 54, 711−721

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Concurrent Cooperativity and Substrate Inhibition in the Epoxidation of Carbamazepine by Cytochrome P450 3A4 Active Site Mutants Inspired by Molecular Dynamics Simulations

Christian S. Müller,[†,‡] Tim Knehans,[§] Dmitri R. Davydov,*[,∥,⊥] Patricia L. Bounds,[‡] Ursula von Mandach,[†] James R. Halpert,[∥] Amedeo Caflisch,*[,§] and Willem H. Koppenol*[,‡]

[†]Department of Obstetrics, University Hospital Zurich, Zürich, Switzerland

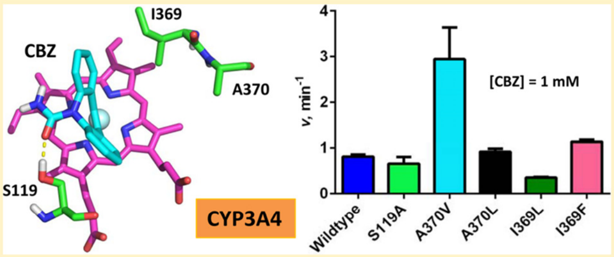
[‡]Institute of Inorganic Chemistry, Swiss Federal Institute of Technology (ETH) Zurich, CH-8093 Zürich, Switzerland

[§]Department of Biochemistry, University of Zurich, CH-8057 Zürich, Switzerland

[∥]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California 92093-0202, United States

[⊥]V. N. Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia

S *Supporting Information*



ABSTRACT: Cytochrome P450 3A4 (CYP3A4) is the major human P450 responsible for the metabolism of carbamazepine (CBZ). To explore the mechanisms of interactions of CYP3A4 with this anticonvulsive drug, we carried out multiple molecular dynamics (MD) simulations, starting with the complex of CYP3A4 manually docked with CBZ. On the basis of these simulations, we engineered CYP3A4 mutants I369F, I369L, A370V, and A370L, in which the productive binding orientation was expected to be stabilized, thus leading to increased turnover of CBZ to the 10,11-epoxide product. In addition, we generated CYP3A4 mutant S119A as a control construct with putative destabilization of the productive binding pose. Evaluation of the kinetics profiles of CBZ epoxidation demonstrate that CYP3A4-containing bacterial membranes (bactosomes) as well as purified CYP3A4 (wild-type and mutants I369L/F) exhibit substrate inhibition in reconstituted systems. In contrast, mutants S119A and A370V/L exhibit S-shaped profiles that are indicative of homotropic cooperativity. MD simulations with two to four CBZ molecules provide evidence that the substrate-binding pocket of CYP3A4 can accommodate more than one molecule of CBZ. Analysis of the kinetics profiles of CBZ metabolism with a model that combines the formalism of the Hill equation with an allowance for substrate inhibition demonstrates that the mechanism of interactions of CBZ with CYP3A4 involves multiple substrate-binding events (most likely three). Despite the retention of the multisite binding mechanism in the mutants, functional manifestations reveal an exquisite sensitivity to even minor structural changes in the binding pocket that are introduced by conservative substitutions such as I369F, I369L, and A370V.

Cytochrome P450 3A4 (CYP3A4) is the most abundant cytochrome P450 enzyme in the liver of most adult humans.[1] This enzyme, which metabolizes about 50% of drugs on the market,[2] is the main hepatic cytochrome P450 involved in the metabolism of 5*H*-dibenzo[*b*,*f*]azepine-5-carboxamide (carbamazepine or CBZ). CBZ is a commonly prescribed anticonvulsive drug; its major metabolite is carbamazepine-10,11-epoxide (Figure 1).[3,4]

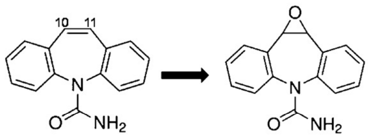


Figure 1. Structures of carbamazepine (left) and the major product of its CYP3A4-dependent oxidation, carbamazepine-10,11-epoxide (right).

Numerous studies on *in vitro* CBZ epoxide formation that include data from incubations with human liver microsomes (HLM),[4−8] recombinant enzyme expressed as microsomes from insect cells,[9,10] and reconstituted purified enzyme preparations[4,11,12] have been published: both hyperbolic and sigmoidal dependencies of the reaction rate on the substrate concentration have been observed. Differing degrees of homotropic cooperativity have been reported for HLM[4,5,7,8] and microsomes from recombinant insect cells expressing human CYP3A4[6,9] that exhibit sigmoidal profiles. Michaelis−Menten kinetics[4] as well as sigmoidal profiles[7] have been reported for purified and reconstituted CYP3A4. The apparent

Received: September 16, 2014
Revised: December 3, 2014
Published: December 29, 2014

inconsistency among these reports demonstrates that epoxidation of CBZ by CYP3A4 is very sensitive to both the experimental conditions and the enzyme source. Such inconsistencies in kinetics profiles have also been reported for other substrates, such as amitriptyline, nifedipine, and testosterone.[13] CBZ qualifies as a narrow therapeutic index drug,[14] of which the target for plasma concentration is 17−70 μM.[7] Because failure to maintain the target plasma range may cause toxic side effects, continuous therapeutic monitoring is desirable for optimal outcome of CBZ therapies. A potential approach to continuous monitoring is to use biosensors constructed from electrodes with immobilized P450 enzymes that can detect drug substrates such as CBZ.[15−17]

In this study, we sought to explore the mechanism of interaction of CYP3A4 with CBZ by engineering CYP3A4 variants with enhanced affinity for CBZ and/or increased epoxidation turnover. These studies were aimed at the elaboration of a CYP3A4-based electrochemical biosensor with improved capacity for determining CBZ concentration in clinical samples. Our approach to the optimization of CYP3A4 for CBZ binding encompasses explicit solvent molecular dynamics (MD) simulations of the structure of the CYP3A4 complex with CBZ in combination with site-specific protein engineering to incorporate amino acid alterations in the enzyme's active site. On the basis of our MD simulations, we designed four CYP3A4 mutants, each with putative stabilization of the CBZ molecule in a predicted productive binding mode that was aimed at increased affinity for CBZ and increased turnover of the epoxide product. We also constructed mutant S119A, which putatively disrupts a critical hydrogen bond, as a negative control with which we expected diminished turnover of the epoxide product.

The chosen mutants were cloned, expressed, and purified, and kinetics parameters of CYP3A4-dependent oxidation of CBZ were studied in a soluble reconstituted system. Significant effects of minor structural changes in the binding pocket of the enzyme, caused by the substitutions S119A, I369L, I369F, A370L, and A370V, were observed, demonstrating a profound sensitivity of both the enzyme−substrate binding interactions and the catalytic efficiency of CYP3A4. The observation of both homotropic cooperativity and substrate inhibition in the kinetics of CBZ epoxidation by wild-type and mutant CYP3A4 suggests that the interactions of the enzyme with the drug involve the binding to the enzyme of at least two, but more likely three, CBZ molecules. This conclusion is supported by multiple MD simulations that demonstrate the possible accommodation of multiple CBZ molecules in the CYP3A4 binding pocket.

■ MATERIALS AND METHODS

Reagents. CBZ was purchased from Acros Organics (Thermo Fisher Scientific, Waltham, MA, USA). L-α-Phosphatidyl choline, L-α-phosphatidyl ethanolamine, and L-α-phosphatidic acid were purchased from Avanti Polar Lipids (Alabaster, AL, USA). 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propansulfonate (CHAPS) was purchased from Applichem GmbH (Darmstadt, Germany). HisPur Ni−NTA beads were obtained from Thermo Fisher Scientific, and CM macroprep ion-exchange support was obtained from Biorad (Hercules, CA, USA). *Escherichia coli* HMS 174 (DE3) cells were purchased from Merck, Darmstadt, Germany. Carbamazepine-10,11-epoxide, DEAE−sepharose fast-flow ion-exchange and adenosine 2′,5′-diphosphate agarose affinity supports were obtained from Sigma-Aldrich (St. Louis, MO, USA). Primers for recombinant DNA manipulations were obtained from Microsynth (Balgach, Switzerland). 10-Methoxy-carbamazepine was obtained from TLC PharmaChem (Vaughan, Ontario, Canada). Bactosomes with a high reductase/CYP3A4 ratio (Human CYP3A4R EasyCYP Bactosomes) were purchased from CYPEX, Ltd. (Dundee, U.K.).

Computational Methods. *Molecular Dynamics Simulations.* The simulations were carried out with the coordinates extracted from the structure of the CYP3A4 complex with metyrapone (Protein Data Bank (PDB): 1W0G).[18] CBZ was superimposed onto the bound structure of metyrapone, and water molecules, ions, and the metyrapone molecule were removed. Because of a lack of crystallographic density in the flexible segments, spurious N and C termini were neutralized with acetyl groups at Asp270 and Ala289 and *N*-methyl-amide groups at Glu262 and Ile276, whereas the natural termini were considered charged. Missing side-chain atoms were added with CHARMM.[19] Parameters for CYP3A4 and HEME were derived from the CHARMM27 force field.[20] The partial charges of the Cys442 side-chain thiolate were set as developed by Foloppe et al.[21] CBZ was parametrized according to the CHARMM general force field[22] and Paramchem.[23,24]

To determine the modes of binding of CBZ to CYP3A4, five independent MD simulations of 50 ns each were carried out with GROMACS (version 4.5.6).[25] First, the CYP3A4−CBZ complex was placed in a dodecahedral box, the size of which was set with a margin of at least 13 Å from any atom of the complex. The box was then filled with pre-equilibrated TIP3P water molecules,[26] and the system was neutralized with Na^+ and Cl^- ions at a concentration of 150 mM. Van der Waals and short-range electrostatic interactions were calculated up to a cutoff of 10 Å. Long-range electrostatics were evaluated with the particle mesh Ewald method,[27] and periodic boundary conditions were applied. All bonds involving hydrogen atoms were constrained by the LINCS algorithm,[28] and a 2 fs time step was used. The temperature was kept constant at 310 K by means of the velocity-rescaling algorithm implemented[29] in GROMACS, and the pressure was kept constant at 1 atm with a Parinello Raman barostat.[30,31] Mutations at the respective residue indices were introduced by simple remodeling in PyMOL.[32] For the simulations with multiple copies of CBZ bound, the starting positions were generated by manually positioning two, three, or four molecules of CBZ in the active site so that no steric clashes were introduced. All systems were energy minimized for 10 000 steps of the conjugate gradient algorithm before starting the MD. Subsequently, the system was equilibrated in an NPT simulation for 1 ns. From the resulting atomic positions, the production runs were started with different random seeds for the initial assignment of the velocities.

Clustering. The WORDOM[33] (version 0.22) implementation of the leader algorithm was used to cluster the snapshots saved along the MD trajectories by a criterion that was based on the root-mean-square deviation (RMSD) of the CBZ nonhydrogen atoms. Only the last 10 ns of each 50 ns run were used for clustering, i.e., the first 40 ns of each run were considered equilibration. First, the CYP3A4 Cα atoms were used to overlap the structures of the coordinate sets in the five 10 ns segments with the X-ray structure. Starting from one of the snapshots at 40 ns, which was taken as the first cluster, the leader algorithm proceeded iteratively by comparing each snapshot with the representative of the previously defined clusters. A snapshot was assigned to a cluster if the RMSD was smaller than a threshold value of 2 Å. Simulations with two CBZ molecules positioned in the active site were clustered with a tree-based algorithm as provided in CAMPARI.[34,35] RMSD-based clustering over the entire 200 ns was carried out for each of the two CBZ molecules separately with a threshold radius (CRADIUS) of 2 Å, a tree height (BIRCHHEIGHT) of 16, and a coarsest threshold (CMAXRAD) of 10 Å. CYP3A4 Cα atoms were aligned before clustering. All figures were prepared with the PyMOL package.[32]

Experimental Section. *Recombinant DNA Manipulations.* Construct pSE3A4, expressing an N-terminally modified human P450 3A4 with a tetra-histidine tag attached at the C-terminus, was described previously.[36] Site-directed mutagenesis was carried out by means of PCR with the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Products Division, La Jolla, CA, USA.).

Expression and Purification of CYP3A4 and P450 Reductase. Freshly transformed *E. coli* Topp3 cells were induced with 0.5 mM IPTG, and the CYP3A4 construct was expressed for 48−72 h before cells were lysed by sonication; expression levels were quantified by reduced CO binding.[37] The enzyme was purified from solubilized membranes via nickel-affinity chromatography on Ni−NTA beads and anion-exchange chromatography with CM macroprep ion-exchange support as described previously.[38,39] Recombinant NADPH cytochrome P450 reductase (CYPOR) from rat liver was expressed in *E. coli* Topp3 cells and was purified as described previously.[36] The purity of the enzymes was analyzed via SDS−polyacrylamide gel electrophoresis and absorbance spectroscopy.

Absorbance Spectroscopy and Analysis of the Spectra. The spectra were recorded at 28 °C with a Specord S 250 UV−vis spectrophotometer equipped with a Peltier thermostat and a stirring device (Analytik Jena, Jena, Germany). The concentrations and the spin state of CYP3A4 were quantified on the basis of the approximation of the heme protein absorbance spectra with a set of the prototype absorbance spectra of CYP3A4's (Fe^{3+}) high-spin, low-spin, and P420 states.[40] The concentration of CYPOR was determined on the basis of the prototypic spectrum of the pure enzyme with an extinction coefficient equal to 21.4 mM^{-1} cm^{-1} at 456 nm.[41] In these calculations, the linear least-squares approximations of the spectra with combinations of the absorbance standards were carried out with the SpectraLab software package.[42]

Reconstitution of Wild-Type Recombinant CYP3A4 and Mutants. A functional mono-oxygenation system with CYP3A4 and CYPOR was reconstituted in a micellar system containing phospholipids added as chloroform solutions of L-α-phosphatidyl choline, L-α-phosphatidyl ethanolamine, and L-α-phosphatidic acid mixed at a molar ratio of 1:2:0.6, respectively,[39] as well as 0.4% CHAPS. The pellet of phospholipids obtained after removal of the solvent under a flow of argon gas was then resuspended in buffer A (100 mM HEPES, 150 mM KCl, 0.5 mM EDTA, and 1 mM DTT, containing 10% *v*/*v* glycerol, pH 7.4) to a final concentration of 10 mg mL^{-1} (13.3 mM) by vigorous shaking on a vortex mixer. The resulting suspension was then homogenized by extrusion through a 200 nm pore-size filter with an Avanti miniextruder (Avanti Polar Lipids, Alabaster, AL, USA.). A mixture of CYP3A4 and CYPOR containing 10 μM of each protein was then supplemented with the 6.5 mM phospholipid suspension to an final concentration of 1.5 mM, and buffer A (1.5% CHAPS) was added to a final concentration of 0.4% CHAPS. The mixtures were incubated overnight at 4 °C under continuous shaking in a table-top thermomixer (Eppendorf, Hamburg, Germany). The reproducibility of the procedure with overnight incubation was considerably better than that with 30 min incubation, and subsequent activity measurements showed no decrease in activity of the reconstituted enzyme after overnight compared to that of after 30 min of shaking.

Kinetics Assay: Formation of Carbamazepine-10,11-epoxide. Reactions were carried out at 37 °C in microreaction tubes in a table-top shaker. The reconstituted enzymes were diluted 20-fold into buffer (0.1 M HEPES, pH 7.4). The final concentrations of CYP3A4 and CYPOR were 0.5 μM, except for the incubations involving CYP3A4 bactosomes, in which the final concentrations were 0.05 μM. The working concentration of CHAPS was 0.02% (325 μM). CBZ was added from methanol stock solutions for a final methanol concentration of 1%, and addition of 20 μL of a 5× NADPH regeneration system containing 25 mM glucose-6-phosphate (GSP), 5 U/mL glucose-6-phosphate dehydrogenase (GSPDH), and 5 mM $NADP^+$ in 50 mM potassium phosphate buffer (pH 7.4) brought the reaction volume to 100 μL and the final concentrations of the regeneration system components to 5 mM, 1 U/mL, and 1 mM, respectively. The reaction mixture was buffered at pH 7.4 with 75 mM HEPES and 10 mM potassium phosphate.

The linear ranges of product formation as a function of time were determined for each mutant by taking aliquots of the reaction mixture at 0, 5, 10, 15, 20, 25, and 30 min after initiation of the reaction. For most of the mutants, we obtained profiles for the increase in product concentration at the time points larger than 20 min that were linear; the exceptions were mutants S119A and I369F, for which the time dependence of product formation was linear for only ca. 15 min. Consequently, incubations were carried out for 20 min for incubations with wild-type CYP3A4 and mutants I369L and A370V/L and for 15 min with mutants S119A and I369F. All incubations were conducted in triplicate. The enzymatic reactions were stopped by adding an equal amount of ice-cold methanol containing 20 μM of the internal standard 10-methoxy-carbamazepine, followed by 1:1 dilution with deionized water for a final methanol concentration of 25%. The samples were analyzed by reversed-phase HPLC on a Ultimate 3000 system (Thermo Fisher Scientific, Reinach, Switzerland) on a 150 mm × 4.5 mm diameter Gravity C18 column (Macherey-Nagel GmBH & Co. KG, Düren, Germany) with 30:70 acetonitrile/water containing 0.1% formic acid as the mobile phase.

The rates of reactions as a function of substrate concentration were fit to either the regular Hill equation or the Hill equation complemented with an allowance for partial substrate inhibition, as described in the Supporting Information, by means of a combination of the Nelder−Mead and Marquardt nonlinear regression algorithms, as implemented in the SpectraLab software package.[42]

■ RESULTS

CBZ Epoxidation by CYP3A4-Containing Bactosomes. In good agreement with previous studies of CYP3A4-dependent metabolism of CBZ, we detected no product in significant amounts except for carbamazepine-10,11-epoxide.[4] Upon examination of the substrate-concentration dependence of the kinetics profiles, we extended the range of CBZ concentrations used in our analyses to 0–2000 μM, a considerably larger concentration range than that reported in previous studies.[4−6,9] The kinetics profile reveals moderate but clearly pronounced substrate inhibition that becomes evident at CBZ concentrations >700 μM (Figure 2). The fit of the initial parts of the titration curves ([CBZ] ≤ 1000 μM) with the Hill equation (Figure 2, dashed line) is consistent with positive cooperativity, with a Hill coefficient (h) of 1.4 ± 0.4 (S_{50} = 162 ± 4 μM, k_{cat} = 1.9 ± 0.3 min^{-1}); this is in agreement with a previous report in which CYP3A4-dependent epoxidation of CBZ was characterized as exhibiting inherent homotropic cooperativity with h = 1.4−1.7.[7]

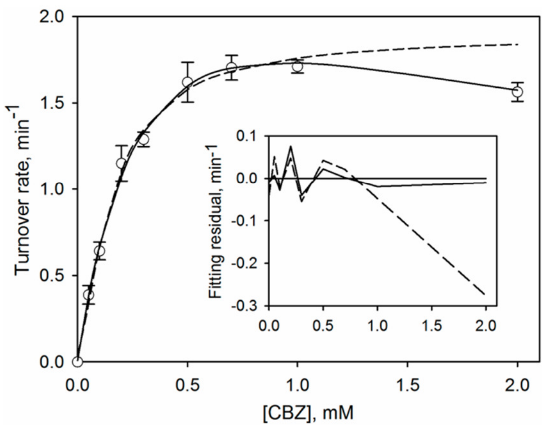


Figure 2. Substrate dependence of the rate of formation of carbamazepine-10,11-epoxide catalyzed by recombinant CYP3A4 in bactosomes. Data points represent mean values from three individual measurements, and the error bars show the respective standard deviations. Results of fitting to the Hill equation (dashed line) and to eq 1 (solid line) are shown. The inset shows plots of the respective residuals.

However, the Hill equation fits of the titration curves are satisfactory only to a CBZ concentration of ca. 700 μM; at higher concentrations, there is evident systematic deviation of the experimental data from the equation curve (Figure 2, inset). In an attempt to achieve an appropriate approximation of the experimental data over the entire range of CBZ concentrations, we fit the data sets to the Michaelis–Menten equation adapted for partial substrate inhibition, expressed as

$$V_S = \frac{k_{cat}[E]_0[S](\alpha[S] + K_I)}{K_M K_I + [S](K_I + [S])} \qquad (1)$$

where V_S and k_{cat} designate the reaction rate at substrate concentration $[S]$ and the apparent catalytic rate constant, respectively. The Michaelis–Menten constant and the dissociation constant of the enzyme complex with the inhibitory substrate molecule are designated as K_M and K_I respectively, and the coefficient α represents the fraction of k_{cat} that is retained in the inhibitory complex when a second molecule of substrate is bound. The derivation of this equation is described in the Supporting Information.

As shown in Figure 2 (solid line), the experimental data sets may be fit to eq 1 ($\rho^2 \geq 0.982$) with no systematic deviations over the entire range of CBZ concentrations studied (Figure 2, inset). The parameters for fitting eq 1, representing the averages of the results obtained in three individual experiments, are k_{cat} = 3.1 ± 0.5 min^{-1}, K_M = 140 ± 10 μM, K_I = 1.6 ± 0.3 mM, and α = 0.21 ± 0.08. Therefore, the data obtained with bactosomes are consistent with a model where the maximal rate of CBZ metabolism is obtained by the complex of CYP3A4 with one substrate molecule, and the subsequent low-affinity binding of a second CBZ molecule to this complex results in an ca. 5-fold decrease in the rate of CBZ metabolism. It appears probable, therefore, that earlier reports of positive cooperativity in CYP3A4-dependent CBZ metabolism should be questioned because the titration curves thought to fit the Hill equation (Figure 2, dashed line) were recorded over a range of CBZ concentrations too low to cause pronounced substrate inhibition.

Probing the CBZ Binding Mode to CYP3A4 with MD Simulations. According to the above analysis, we infer that the complex of CYP3A4 with one substrate molecule in the substrate-binding pocket may serve as an adequate initial model of the catalytically competent complex of the enzyme with CBZ. Thus, we selected a binary complex of CYP3A4 with one CBZ molecule as the subject for initial MD simulations aimed at eliciting the structure of the CBZ–CYP3A4 complex and the orientation of the substrate molecule in the active site.

Figure 3b shows the time series of the backbone RMSD from the X-ray structure of CYP3A4 (PDB: 1W0G),[18] which reaches a plateau between 2 and 3 Å after about 10 ns (equilibration phase). We determined the binding modes of CBZ in five independent MD runs of 50 ns each, where only the last 10 ns of each run was evaluated. In the most populated, i.e., largest, cluster of the CBZ poses, the heme iron is located within 5 Å of the center of the bond between the carbon atoms 10 and 11 (C10,C11), which is compatible with the CYP3A4-catalyzed epoxidation of CBZ (Figure 3a). It should be noted, however, that the relevance of this binding mode to the actual mechanism of interactions is difficult to assess because it was primarily sampled in only one of five MD runs. The CBZ molecule that was initially superimposed at the position of the removed metyrapone did not exit the CYP3A4 binding site in any of the runs. On the basis of the binding mode of CBZ in the largest cluster of poses deduced from MD simulations (Figure 3a), we selected the amino acids S119, A370, and I369 as targets (Table S1). We suggested five single-point mutants on the basis of the mean distances between the center of mass (COM) of the respective residues and the COM of CBZ in this

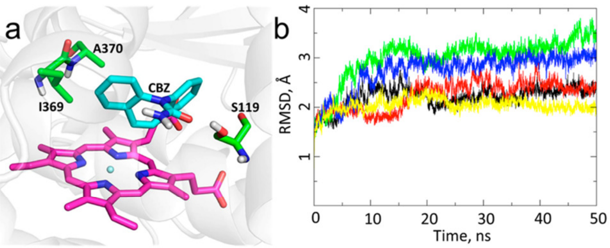


Figure 3. Results of MD simulations with the complex of wild-type CYP3A4 with one molecule of CBZ. (a) A representative snapshot of the most populated cluster. The residues selected for mutation (S119, I369, and A370), CBZ, and the heme group are shown with stick models in green, cyan, and magenta, respectively. The protein backbone is shown in transparent gray. (b) Time series of the RMSD from the crystal structure (PDB: 1W0G) of the protein backbone after alignment of Cα carbons. The RMSD reaches a plateau after about 20 ns in the five independent runs, each of which is plotted with a different color.

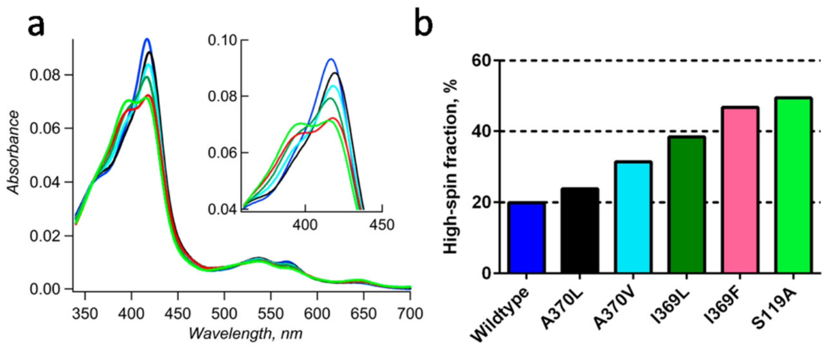


Figure 4. Effect of mutations on the spin state of purified CYP3A4 proteins. (a) Absorbance spectra of wild-type CYP3A4 (blue) and its mutants: I369F (red), I369L (green), S119A (light green), A370V (turquoise), and A370L (black). Spectra were normalized to correspond to a heme protein concentration of 1 μM. The inset shows the Soret band region of the same spectra. (b) Respective fraction of the high spin state of these heme proteins. Conditions: 100 mM HEPES, pH 7.4, 1 mM DTT, 10% *v*/*v* glycerol. The concentration of P450 was in the range of 1.5−2.5 μM.

cluster (Figure 3a). We predicted that mutants I369L, I369F, A370V, and A370L would lead to improved steric complementarity, i.e., Van der Waals interactions, between CYP3A4 and the 5*H*-dibenzo[*b*,*f*]azepine ring system of CBZ. Mutant S119A was selected as a negative control and was designed to disrupt the structural stability of the productive binding mode by eliminating the hydrogen bond between the carbonyl oxygen of CBZ and the S119 side-chain hydroxyl group.

Enzyme Expression, Purification, and Reconstitutions. On the basis of the results of the MD simulations, we elected to probe the effect of the mutations S119A, I369L, I369F, A370L, and A370V on the metabolism of CBZ by CYP3A4. Expression of all five constructs in *E. coli* provided stable P450 holoprotein, which is in good agreement with earlier reports on the expression and purification of mutants S119A,[43] A370V,[44] and I369F.[45] The yields of expression obtained with all of these mutants were generally somewhat lower than the yield of wild-type CYP3A4.

As seen from the absorbance spectra of the purified enzymes (Figure 4a), all five substitutions probed in this study caused a notable increase in the amplitude of the heme protein absorbance band at 396 nm relative to the amplitude of the band at 417 nm. This is indicative of a displacement of the spin equilibrium toward the ferric high-spin state. This displacement is most pronounced in mutants I369F and S119A, where the fraction of high-spin heme protein approaches 50%, in sharp contrast to the spin state of wild-type CYP3A4, where the high-spin fraction at 28 °C does not exceed 20% (Figure 4b).

Effect of Mutations on CYP3A4-Dependent Epoxidation of CBZ. In initial experiments, we assessed the turnover of CBZ epoxidation in the reconstituted system with the purified mutants and wild-type CYP3A4 (rCYP3A4) at apparent subsaturating (0.1 mM) and saturating (1 mM) concentrations of the substrate.

To analyze the effects of these substitutions in more detail, we evaluated broader kinetics profiles of wild-type CYP3A4 and mutants in reconstituted system, increasing the range of CBZ concentration studied up to 2 mM (Figure 5, Table 1). Although the overall quality of the fit of the titration curves obtained with wild-type CYP3A4 to the Hill equation (Figure 5, dashed line, and Table 1, $h = 2.0 \pm 0.9$, $S_{50} = 300 \pm 100\ \mu$M, and $k_{cat} = 1.0 \pm 0.3\ \text{min}^{-1}$) appears acceptable ($\rho^2 = 0.957$), systematic deviations of the experimental points (Figure 5a, circles) suggest the presence of substrate inhibition at higher concentrations, similar to that observed with the CYP3A4-containing bactosomes. However, in the case of mutants I369F and I369L, the substrate inhibition is much more evident than with wild-type CYP3A4 (Figure 5a); mutants S119A, A370V, and A370L show no visible signs of substrate inhibition (Figure 5b). In contrast, the S-shaped kinetics profiles obtained with these mutants are indicative of a high degree of cooperativity.

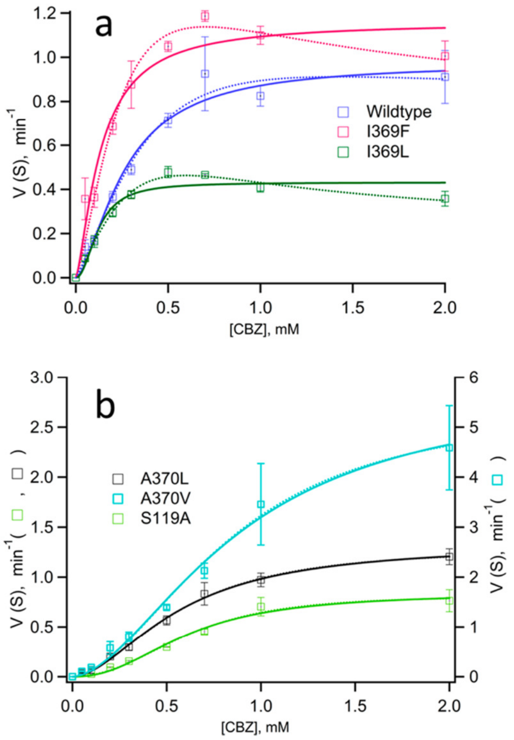


Figure 5. Substrate dependence of the rate of formation of carbamazepine-10,11-epoxide catalyzed by purified CYP3A4 and mutants in the reconstituted system. Incubations were carried out with 0.5 μM CYP3A4. Data points represent the mean values of three individual measurements, and the error bars show the respective standard deviations. Fits to the Hill equation (dashed line) and to eq 2 (solid line) are shown. (a) Plots obtained with wild-type CYP3A4 (circles) and the mutant enzymes I369F (squares) and I369L (triangles). (b) Plots obtained with mutants S119A (circles), A370V (squares), and A370L (triangles).

To justify the apparent combination of the positive cooperativity and substrate inhibition mechanisms observed

Table 1. Parameters of CYP3A4-Dependent Epoxidation of Carbamazepine[a]

	Hill equation[b]				Hill equation complemented with substrate inhibition (eq 2)					
system	k_{cat}, min^{-1}	S_{50}, mM	h	ρ^2	k_{cat}, min^{-1}	S_{50}, mM	h	K_i, mM	α	ρ^2
bactosomes	1.9 ± 0.3	0.16 ± 0.04	1.4 ± 0.4	0.979	2.9 ± 0.7	0.3 ± 0.2	1.06 ± 0.08	1.3 ± 0.8	0.3 ± 0.3	0.991
rCYP3A4 w/t	0.9 ± 0.3	0.3 ± 0.1	2.0 ± 0.9	0.957	1.4 ± 0.4	0.50 ± 0.04	1.4 ± 0.4	0.79 ± 0.04	0.6 ± 0.5	0.964
rCYP3A4 I369L	0.42 ± 0.06	0.14 ± 0.02	1.80 ± 0.10	0.884	0.8 ± 0.4	0.3 ± 0.2	1.4 ± 0.3	1.1 ± 0.5	0.2 ± 0.1	0.989
rCYP3A4 I369F	1.4 ± 0.2	0.20 ± 0.07	1.3 ± 0.5	0.932	3 ± 1	0.6 ± 0.4	1.2 ± 0.3	1.1 ± 0.6	0.2 ± 0.4	0.971
rCYP3A4 S119A	0.8 ± 0.1	0.61 ± 0.07	2.5 ± 0.4	0.985	1.2 ± 0.1	0.94 ± 0.08	2.1 ± 0.3	1.0 ± 0.2	0.5 ± 0.1	0.990
rCYP3A4 A370V	5 ± 1	1.0 ± 0.3	2.1 ± 0.9	0.991	4.6 ± 0.9	1.1 ± 0.3	1.9 ± 0.6	5 ± 8	0.96 ± 0.07	0.991
rCYP3A4 A370L	1.3 ± 0.4	0.6 ± 0.3	2.0 ± 0.6	0.993	1.2 ± 0.2	0.77 ± 0.07	1.7 ± 0.2	5 ± 9	1.4 ± 0.7	0.993

[a] The values given in the table represent the averages of three individual measurements, and the values after ± show the confidence intervals calculated for $p = 0.05$. [b] In cases with clearly pronounced substrate inhibition (CYP3A4 bactosomes, reconstituted systems with mutants I369L and I369F), the fitting to the conventional form of the Hill equation was carried out with the data subsets corresponding to a CBZ concentration range of 0−1 mM. The square correlation coefficients given in the table correspond to the entire region of CBZ concentrations (0−2 mM) in all cases.

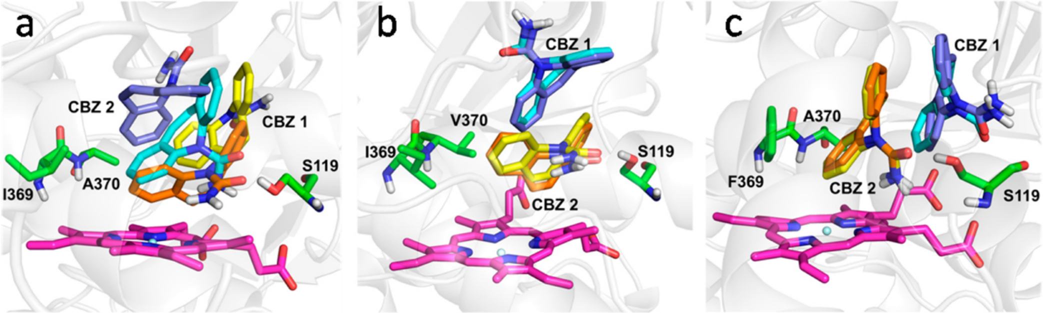


Figure 6. Results of MD simulations with the complex of two molecules of CBZ and (a) wild-type CYP3A4, (b) A370V, and (c) I369F. Each panel shows two representatives that correspond to the most populated clusters obtained with each of the two CBZ molecules for clustering. Clustering of the non-hydrogen atoms of the CBZ molecules was carried out upon structural alignment of all snapshots of the MD trajectories to the X-ray structure via the Cα atoms of the α-helical residues of CYP3A4. Only in panel a are the orientations of the CBZ molecules different, depending on which of the two CBZ molecules was subjected to clustering, indicating that the sampling did not converge.

in the mutants, we attempted to approximate the experimental data with an equation representing a combination of the Hill equation formalism with an allowance for substrate inhibition, represented as

$$V_S = \frac{k_{cat}\cdot[E]_0\cdot[S]^h\cdot(K_I + \alpha\cdot[S])}{S_{50}^h\cdot K_I + [S]^h\cdot(K_I + [S])} \qquad (2)$$

The derivation of eq 2 is described in the Supporting Information. The S_{50} term in eq 2 is defined in a way similar to that in the regular Hill equation: as the hth root of the apparent dissociation constant of the catalytically competent enzyme−substrate complex. Where $h = 1$, eq 2 reduces to eq 1, and S_{50} is equivalent to K_M.

The fits of eq 2 to the data are shown in Figure 5 (solid lines), and the respective parameters are summarized in Table 1. Equation 2 clearly provides a better model for fitting data sets obtained with wild-type CYP3A4, whether prepared as bactosomes or as reconstituted systems, and with mutants I369L and I369F. For the data obtained with the mutants, the fit to eq 2 is as good as that to eq 1. The values obtained for the parameter α with mutants A370L and A370V are close to unity, which suggests that the binding of additional substrate molecules would not affect the activity of the enzyme; thus, we conclude that substrate inhibition is completely abolished in these mutants.

Analysis of the kinetics parameters suggests that, over the five mutations probed in this study, only the I369F substitution and, especially, the A370V substitution cause considerable increases in enzyme turnover with CBZ compared with wild-type (Table 1). In the case of I369F, the increase in the k_{cat} value is attenuated by enhanced substrate inhibition that is caused by a decrease in the fraction of activity retained in the inhibitory complex (α). However, for the A370V mutant, where substrate inhibition is eliminated, the increase in the CBZ turnover is substantial (Figure 5b, Table 1).

The clearly pronounced positive cooperativity in CBZ metabolism observed with mutants A370V, A370L, and S119A suggests that the formation of the catalytically competent complex with CBZ in these mutants requires the binding of at least two substrate molecules per molecule of the enzyme. At the same time, in contrast to mutants A370V and A370L, where substrate inhibition is eliminated, the fit of the data obtained with S119A to eq 2 indicates that the metabolism of CBZ by this mutant is characterized by a combination of a prominent positive cooperativity ($h = 2.1 \pm 0.3$) with a moderate degree of substrate inhibition ($\alpha = 0.5 \pm 0.1$); this further suggests that the mechanism of interaction with CBZ involves at least three separate substrate binding events.

MD Simulations of Two to Four CBZ Molecules in the CYP3A4 Binding Pocket. On the basis of the analysis of the kinetics profiles for CBZ metabolism, we concluded that changes in CBZ metabolism caused by the mutations at positions 369, 370, and 119 cannot be accurately described by simulations with a single CBZ molecule bound. Therefore, we conducted MD simulations with two, three, or four CBZ molecules positioned in the CYP3A4 binding pocket. Five MD simulations each with two copies of CBZ in the active site were carried out with wild-type CYP3A4 and mutants I369F and A370V. Cluster analyses yielded nonidentical largest clusters for both mutants and wild-type. The simulation times of 0.2 μs per MD run, equal to a total sampling time of 1 μs for each of the three CYP3A4 variants, were intended to achieve convergent data. Importantly, similar to the single-CBZ simulations, the clusters obtained stem mainly from only one of the five runs, which is indicative of a lack of convergence. The largest clusters for each of wild-type CYP3A4 and mutants A370V and I369F are shown in Figure 6. Only for the A370V mutant (Figure 6b) is the binding mode, i.e., the position and orientation, of the CBZ closest to the heme essentially identical to that of the most populated cluster in the single-CBZ wild-type run with the C10−C11 bond close to the heme iron (Figure 3a). Interestingly, the hydrogen bond observed between S119 and the carbonyl oxygen of one CBZ molecule is retained in all panels in Figure 6, which further supports our hypothesis that S119 plays an important role in substrate binding.

Partial dissociation of a single CBZ molecule was observed in one of the multiple-CBZ MD simulations (Figures 7 and S1). Analysis of the unbinding pathway reveals that the CBZ molecule moves to the periphery of CYP3A4 through the F′-helix β-sheet-1 putative substrate-binding channel predicted by Williams et al.[18] To assess the ability of CYP3A4 to accommodate more than two CBZ molecules, we carried out five independent MD runs of 100 ns each with wild-type CYP3A4 plus three or four CBZ molecules (Figure S1). No dissociation events were observed, and the distance of the CBZ molecule furthest from the heme iron was never larger than 18 Å. These results indicate that the CYP3A4 binding pocket is large enough to accommodate up to four CBZ molecules and that the A370V and I369F mutations considerably affect the mobility of CBZ molecules bound in the binding pocket.

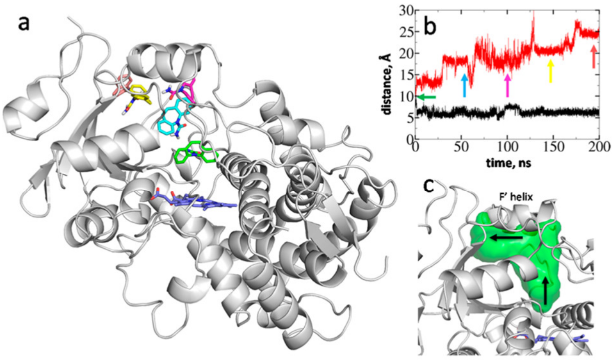


Figure 7. Pathway of egress from the heme pocket for one of the two bound CBZ molecules from wild-type CYP3A4 to a distance >20 Å from the heme iron. (a) Structure of CYP3A4 showing the positions of the CBZ molecule during its dissociation. The initial position of the CBZ molecule is highlighted in green and the positions after 50, 100, 150, and 200 ns are shown in cyan, magenta, yellow, and maroon, respectively. The second CBZ molecule is not shown. (b) Time series of the COM distance between CBZ C10,C11 and the heme iron. The distances of the bound CBZ molecule are shown in black, and the distances of the unbinding CBZ molecule are shown in colors corresponding to those used in panel a. The green arrow indicates the starting position of the molecule. (c) Surface circumscribed by the central nitrogen atom of the dissociating CBZ molecule over the course of a single 200 ns simulation. The protein backbone from the first snapshot of the simulation is shown in white and does not account for changes during the simulation. The arrows indicate the general direction of the movement of the CBZ molecule.

■ DISCUSSION

The present study represents the first attempt to employ MD simulations to produce CYP3A4 mutants with enhanced activity of epoxidation of carbamazepine combined with subsequent *in vitro* examination of the impact of the mutations on kinetics parameters. Our initial studies of the kinetics of CBZ epoxidation by CYP3A4-containing bactosomes reveal pronounced substrate inhibition, leading us to suggest that the apparent positive cooperativity reported in earlier studies of CBZ metabolism might be due to the limited range of substrate concentrations examined. We demonstrate that the kinetics profiles obtained with bactosomes adequately approximate the Michaelis−Menten equation complemented with an allowance for partial substrate inhibition (eq 1). The complex of CYP3A4 with one substrate molecule bound in the substrate-binding pocket (Figure 3) may be taken as a reasonable model of the catalytically competent complex.

In the MD simulations carried out with a single CBZ molecule, the most populated cluster displays a productive CBZ binding mode that is compatible with the epoxidation reaction. The distance between the COM of CBZ carbon atoms 10 and 11 to the heme iron is less than 5 Å. On the basis of the analysis of the structure in this cluster, we identified CYP3A4 mutations (I369L, I369F, A370V, and A370L) that could lead to improved steric complementarity via Van der Waals interactions of the enzyme with the three-ringed 5*H*-dibenzo-[*b*,*f*]azepine system of CBZ. The S119A mutation was designed to abolish the hydrogen bond to the carboxyl function of CBZ to decrease the stability of the binding mode and, thus, the overall binding affinity.

Although the approach that was based on MD simulations with one CBZ molecule was successful in the case of mutants I369F and A370V, where the maximal rate of CBZ turnover is considerably increased, examination of the kinetics profiles of CBZ metabolism with the full series of mutants reveals certain limitations of this simplified model. We can best approximate the kinetics profiles obtained *in vitro* with wild-type CYP3A4 or mutants by means of a model that combines the positive cooperativity mechanism with an allowance for partial substrate inhibition (eq 2). According to our analysis with this model, marginal cooperativity (h = 1.2−1.4) is exhibited by wild-type CYP3A4 and mutants I369L and I369F, whereas the cooperativity is more clearly pronounced (h = 1.7−2.1) in mutants S119A, A370V, and A370L. Substrate inhibition is quite modest in wild-type CYP3A4 and the S119A mutant ($\alpha \approx$ 0.5) and eliminated in mutants A370V and A370L ($\alpha \approx 1$), but it is considerably enhanced in mutants I369L and I369F ($\alpha \approx$ 0.2). The complex mix of positive cooperativity and substrate inhibition in CBZ metabolism are consistent with the results of Egnell and co-workers,[7] who compared the kinetics profiles for CBZ epoxidation by HLM to those obtained with purified CYP3A4 in a reconstituted system and demonstrated that although data obtained with HLM are consistent with a two-site model where binding of the first CBZ molecule increases the enzyme affinity for binding of a second substrate molecule the reconstituted system instead exhibits substrate inhibition.[7]

On the basis of our findings, we suggest that CYP3A4−CBZ interaction is best explained with a model that involves binding of three or more CBZ molecules in the active site. MD simulations of the complex of CYP3A4 with two to four CBZ molecules provide strong support for this suggestion; no complete egress of substrate was observed in five MD simulations involving three or four CBZ molecules. In the simulations where two molecules of CBZ are bound to wild-type CYP3A4 or mutants I369F or A370V (Figure 6), the most populated binding mode of the single-copy simulations of CBZ was observed only for A370V: in this case, one of the two CBZ molecules is bound in a putative catalytic pose. The finding that this pose appears to be stabilized by the second CBZ molecule provides a plausible explanation for homotropic cooperativity in CBZ epoxidation. Indeed, the A370V mutant is the only one of the three enzyme variants probed with simulations that exhibits prominent homotropic cooperativity (Figure 5, Table 1). However, given the limited sampling in the simulations, we caution that this observation may be coincidental.

In the 15 simulations of 2 CBZ molecules in the active site of CYP3A4 (5 simulations each with wild-type, A370V, and I369F), only one partial egress of substrate was observed: the case of wild-type CYP3A4 (Figure 7). We compared the cumulative volume of the dissociation trajectory of the CBZ undergoing partial dissociation via the sampled volume of the COM of C10/11 (Figure 7c) to the position of the substrate access channel predicted by Williams et al. on the basis of the X-ray structure (PDB: 1W0G) of CYP3A4[18] and found that the volumes largely match; hence, our simulation supports the geometry of the substrate- and solvent-accessible binding pocket of CYP3A4 previously proposed.

Although a multisite binding mechanism in mutants is retained, the enzyme is exquisitely sensitive to the minor structural changes in the binding pocket introduced by the point mutations S119A, I369F, I369L, A370L, and A370V. The observation of the hydrogen bond between the CBZ molecule adjacent to the heme and the S119 residue (Figure 6) provides a clue to the effects of the S119A substitution (Figure 4, Figure 5, Table 1). Importantly, this bond is found in the structures of all simulated variants (wild-type, A370V, and I369F). The absence of the stabilizing hydrogen bond in the S119A mutant may explain the reduced affinity of this mutant for CBZ. Further, enhancement of homotropic cooperativity in CBZ epoxidation by this mutation reveals an increased role for the binding of a second substrate molecule in stabilizing the catalytically competent complex. These results also support that residue S119 plays a key role in substrate binding in CYP3A4,[44,46−50] in agreement with a recent study in which it was shown that the association of CYP3A4 with ritonavir analogues, which do not engage in hydrogen bonding to S119, is facilitated by polar interactions mediated by the S119 residue.[50]

Possible effects of the mutations on the structure of CYP3A4 are illustrated in Figure 8, which shows the largest clusters obtained in simulations with two CBZ molecules of wild-type CYP3A4 and mutants I369F and A370V compared to the X-ray crystal structure of CYP3A4 with metyrapone (PDB: 1W0G).[18] The $\pi-\pi$ stacking interactions between residues F316 and F367 that are also found in most X-ray structures are likely to be structurally important. The residues I369 and A370 reside in the 1.4 K β loop (residues 368−372) that connects the α helix K (violet) with the subsequent β strand (blue) and adjoins the active site in Van der Waals contact to the α helix I (green). The structural changes upon substitution of phenylalanine at I369 are apparently important for the increased substrate inhibition observed in the I369F mutant (Figure 5, Table 1). The substitution at I369 with phenylalanine (or leucine) is likely to affect hydrophobic interactions in the active site, thereby promoting inhibitory substrate-binding and/or steric or conformational effects. The I369F substitution may reduce the volume of the active site and concomitantly restrict access to the heme (Figure 8c). The bulky phenylalanine residue 369 in the vicinity of the heme is likely to destabilize the ligation of the water molecule in the low-spin P450 state and to be the source of the observed increase in the proportion of high-spin P450 in the substrate-free I369F mutant (Figure 4). The I369F substitution may reduce the empty space between the 1.4 K/β loop and helix I that face each other in the active site. An analysis of the trajectories of residues F304 and 369 in the MD simulations (Figure S2) reveals a decrease of the mean distance between the side chains of these residues (blue dashes in Figure 8b−d) from ∼11.6 Å in wild-type CYP3A4 ($\Delta_{SC(WT)}$, Figures 8b and S2) to ∼8.0 Å in I369F ($\Delta_{SC(I369F)}$, Figures 8c and S2)

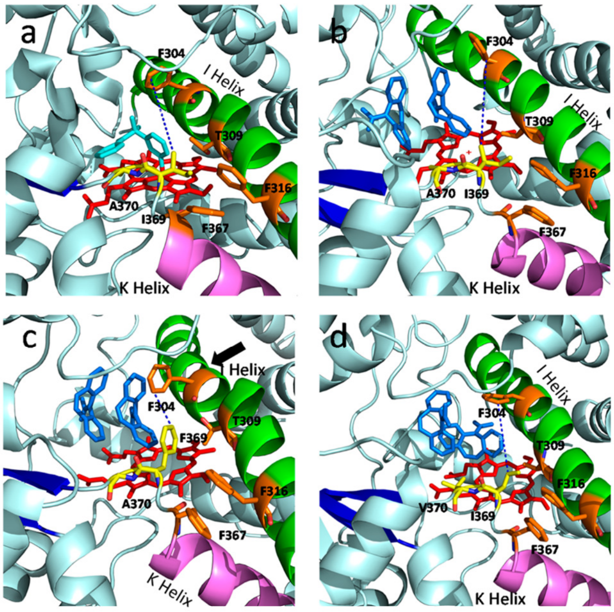


Figure 8. Comparison of the X-ray structure of CYP3A4 (PDB: 1W0G) with the simulated structure cluster representatives. (a) Active site of CYP3A4 structure with metyrapone (light blue) bound.[18] (b−d) Cluster representatives of MD simulations of (b) CYP3A4 wild-type and mutants (c) I369F and (d) A370V with two CBZ molecules (marine blue). Residues F316 and F367 (orange) apparently engage in $\pi-\pi$ stacking interactions. The 1.4 K/β loop (consisting of residues 368−372 (PIAMR) that join the K helix (violet) to the sequential β sheet (blue)) makes Van der Waals contacts with the long I helix (green). Mutated residues are shown in yellow. Note: in panel c, the bulky phenylalanine in the I369F mutant projects into the active site, reducing the accessibility of the heme. Over the length of one simulation of the I369F mutant, the I helix exhibits a kink (black arrow) around residue F304 that additionally restricts access to the active site because of a dramatic decrease in the distance between the side-chains of F304 and residue 369 (blue dashes). The average distances over the total simulation time are (b) $\Delta_{SC(WT)} \approx 11.6$ Å in wild-type CYP3A4, (c) $\Delta_{SC(I369F)} \approx 8.0$ Å in the I369F mutant, and (d) $\Delta_{SC(A370V)} \approx 13.0$ Å in the A370V mutant.

on average over the total simulation; the corresponding distance in the A370V mutant was ~13.0 Å ($\Delta_{SC(A370V)}$, Figures 8d and S2). The formation of a kink in the helix near the F304 residue (Figure 8c, black arrow) may further restrict access to the active site in the I369F mutant: this conformational change was, however, seen in only one out of five simulations. Furthermore, although not revealed in the simulations, a direct interaction of the I369F phenylalanine with residue F304 may be possible and contribute to the changes in the kinetics observed with this mutant. In the I369L mutant, the side-chain γ-methyl groups of leucine may more closely interact with the T309 side chain to move the 1.4 K/β loop and to induce a similar conformational change in helix I as seen in the simulation of the I369F mutant. A direct effect of the leucine on the accessibility to the heme may also be plausible.

Stabilization of substrate binding by the additional phenylalanine residue near the heme moiety in the I369F mutant via stacking interactions of the aromatic rings may contribute to the high rate of CBZ turnover observed with this mutant at low CBZ concentrations. At higher CBZ concentrations, the presence of several CBZ molecules stacked on top of each other above the heme may prevent dissociation of product and be responsible for the observed substrate inhibition (Table 1).

Elimination of substrate inhibition in mutants A370V and A370L (Figure 5, Table 1) is associated with a significant increase in the S_{50} value for these mutants as compared to that of wild-type, and the decreased affinity for CBZ provides further support for the importance of the K/β-loop region for CBZ binding. It seems plausible that the steric clashes between the larger aliphatic side chains of the valine and leucine of these mutants and the metabolized or stabilizing CBZ molecule may destabilize the substrate stacking, which is hypothesized to be responsible for substrate inhibition.

The observation that the mutants with higher affinities for CBZ display more pronounced substrate inhibition provides additional support for the above hypothesis that the mechanism of substrate inhibition is due to the binding of additional CBZ molecules that impede the egress of product. In contrast, the finding that the mutants with lower affinities for CBZ exhibit homotropic cooperativity is consistent with the stabilization of the catalytically competent enzyme−substrate complex upon binding of a second CBZ. Whether these correlations are specific to metabolism of CBZ or represent a general characteristic of CYP3A4-catalyzed reactions requires further experimentation.

In view of our original goal to generate CYP3A4 variants for use in biosensors capable of detecting CBZ, one may argue that the A370V variant is the best candidate because it exhibits the highest maximal rate of CBZ turnover of all the probed enzyme variants. However, taking into account that the target therapeutic concentrations of CBZ in blood plasma are limited to the range of 17 to 70 μM,[7] we may conclude that the decreased affinity of A370V for CBZ (S_{50} = 1.1 mM) undermines its practical utility. In our view, the most promising enzyme variant for practical use is I369F, where the rate of CBZ turnover is increased 2-fold and the S_{50} value obtained with the use of eq 2 remains unaffected (Table 1). However, it should be taken into account that the rate of substrate turnover and the utilization of the reducing equivalents by the enzyme do not necessarily obey a strict proportionality. Baas et al. showed that the consumption of NADPH as a source of reducing equivalents for CYP3A4 reaches its maximum at substrate concentrations that are subsaturating for the enzymatic turnover of testosterone.[51] Furthermore, the actual relationships between the faradaic current to the electrodes and the enzyme turnover may differ between the mutants. Therefore, further assessment of the potential utility of the generated CYP3A4 variants for the use in CBZ sensors requires investigations with the enzyme immobilized on the electrode surface.

Determination of the crystal structures of the CYP3A4 variants investigated in this study, such as I369F and S119A, would be expected to yield important insights regarding the relevance of structural aspects to mechanisms of substrate inhibition and homotropic cooperativity in CYP3A4. In spite of the structural plasticity of CYP3A4,[47,52] we have been able to use MD simulations to identify plausible productive and nonproductive binding modes for one or more substrate molecules in the substrate-binding pocket of the enzyme. In particular, the combination of MD simulations and pose clustering to guide site-directed mutagenesis has resulted in the design of enzyme variants that exhibit altered functional properties, such as homotropic cooperativity, substrate inhibition, and substrate affinities, *in vitro*. This approach may prove to be generally applicable to other enzyme−substrate pairs as a means to probe binding interactions and mechanisms and to modulate functionality.

■ ASSOCIATED CONTENT

S Supporting Information

Information regarding selection and suggestion of mutants, COM tables, equation derivation steps, distances of the CBZ C10,11 COM to the heme iron in MD simulations, and time series data of the distance between the center of the C10−C11 bond in CBZ and the heme iron and between the side chains of residue 369 and Phe304. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*D.R.D.: Telephone: +1-559-751-2309. E-mail: dmrdavyd@gmail.com.
*A.C.: Telephone: +41-44-635-5521. E-mail: caflisch@bioc.uzh.ch.
*W.H.K.: Telephone: +41-44-632-2875. E-mail: koppenol@inorg.chem.ethz.ch.

Present Addresses

P.L.B.: Foundation for Research on Information Technologies in Society (IT'IS), Zeughausstrasse 43, CH-8004 Zurich, Switzerland.
J.R.H.: School of Pharmacy, University of Connecticut, Storrs, Connecticut 06269 United States.

Funding

This research was supported in part by grant GM054995 from the National Institutes of Health (J.R.H.), grant 315230_149897 from the Swiss National Science Foundation (A.C.), Sinergia grant 127547 of the Swiss National Science Foundation (U.M), and the Swiss Federal Institute of Technology (W.H.K.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Michael Fairhead, Michael Richter, and Linda Thöny for their valued support of C.M. in experimental molecular biology techniques. We thank Marino Convertino for help with the simulation setup and protocols in the initial phase of this project. We thank Antonio Baici for instrumental comments on initial interpretation of the kinetics. C.M. thanks Alexandra Dolder for technical assistance in preparation and analysis of the proteins.

■ ABBREVIATIONS

CBZ, carbamazepine (5*H*-dibenzo[*b*,*f*]azepine-5-carboxamide); CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; COM, center of mass; CYP3A4, cytochrome P450 3A4; CYPOR, cytochrome P450 oxidoreductase; MD, molecular dynamics; rCYP3A4, reconstituted (recombinant) CYP3A4

■ REFERENCES

(1) Shimada, T., Yamazaki, H., Mimura, M., Inui, Y., and Guengerich, F. P. (1994) Interindividual variants in human liver cytochrome-P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver-microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther. 270*, 414−423.

(2) Guengerich, F. P. (1999) Cytochrome P-450 3A4: Regulation and role in drug metabolism. *Annu. Rev. Pharmacol. Toxicol. 39*, 1−17.

(3) Pearce, R. E., Vakkalagadda, G. R., and Leeder, J. S. (2002) Pathways of carbamazepine bioactivation *in vitro* I. Characterization of human cytochromes P450 responsible for the formation of 2- and 3-hydroxylated metabolites. *Drug Metab. Dispos. 30*, 1170−1179.

(4) Kerr, B. M., Thummel, K. E., Wurden, C. J., Klein, S. M., Kroetz, D. L., Gonzalez, F. J., and Levy, R. (1994) Human liver carbamazepine metabolism: Role of CYP3A4 and CYP2C8 in 10,11-epoxide formation. *Biochem. Pharmacol. (Amsterdam, Neth.) 47*, 1969−1979.

(5) Nakamura, H., Nakasa, H., Ishii, I., Ariyoshi, N., Igarashi, T., Ohmori, S., and Kitada, M. (2002) Effects of endogenous steroids on CYP3A4-mediated drug metabolism by human liver microsomes. *Drug Metab. Dispos. 30*, 534−540.

(6) Nakamura, H., Torimoto, N., Ishii, I., Ariyoshi, N., Nakasa, H., Ohmori, S., and Kitada, M. (2003) CYP3A4 and CYP3A7-mediated carbamazepine 10,11-epoxidation are activated by differential endogenous steroids. *Drug Metab. Dispos. 31*, 432−438.

(7) Egnell, A.-C., Houston, B., and Boyer, S. (2003) In vivo CYP3A4 heteroactivation is a possible mechanism for the drug interaction between felbamate and carbamazepine. *J. Pharmacol. Exp. Ther. 305*, 1251−1262.

(8) Egnell, A.-C., Houston, J. B., and Boyer, C. S. (2005) Predictive models of CYP3A4 heteroactivation: *In vitro−in vivo* scaling and pharmacophore modeling. *J. Pharmacol. Exp. Ther. 312*, 926−937.

(9) Maekawa, K., Yoshimura, T., Saito, Y., Fujimura, Y., Aohara, F., Emoto, C., Iwasaki, K., Hanioka, N., Narimatsu, S., Niwa, T., and Sawada, J. (2009) Functional characterization of CYP3A4.16: Catalytic activities toward midazolam and carbamazepine. *Xenobiotica 39*, 140−147.

(10) Torimoto, N., Ishii, I., Hata, M., Nakamura, H., Imada, H., Ariyoshi, N., Ohmori, S., Igarashi, T., and Kitada, M. (2003) Direct interaction between substrates and endogenous steroids in the active site may change the activity of cytochrome P450 3A4. *Biochemistry 42*, 15068−15077.

(11) Ueng, Y.-F., Kuwabara, T., Chun, Y.-J., and Guengerich, F. P. (1997) Cooperativity in oxidations catalyzed by cytochrome P450 3A4. *Biochemistry 36*, 370−381.

(12) Korzekwa, K. R., Krishnamachary, N., Shou, M., Ogai, A., Parise, R. A., Rettie, A. E., Gonzalez, F. J., and Tracy, T. S. (1998) Evaluation of atypical cytochrome P450 kinetics with two-substrate models: Evidence that multiple substrates can simultaneously bind to cytochrome P450 active sites. *Biochemistry 37*, 4137−4147.

(13) Houston, J. B., and Kenworthy, K. E. (2000) *In vitro−in vivo* scaling of CYP kinetic data not consistent with the classical Michaelis−Menten model. *Drug Metab. Dispos. 28*, 246−254.

(14) Yacobi, A., Zlotnick, S., Colaizzi, J. L., Moros, D., Masson, E., Abolfathi, Z., LeBel, M., Mehta, R., Golander, Y., and Levitt, B. (1999) A multiple-dose safety and bioequivalence study of a narrow therapeutic index drug: A case for carbamazepine. *Clin. Pharmacol. Ther. (N.Y., NY, U.S.) 65*, 389−394.

(15) Bistolas, N., Wollenberger, U., Jung, C., and Scheller, F. W. (2005) Cytochrome p450 biosensors−a review. *Biosens. Bioelectron. 20*, 2408−2423.

(16) Dodhia, V. R., Sassone, C., Fantuzzi, A., Nardo, G. D., Sadeghi, S. J., and Gilardi, G. (2008) Modulating the coupling efficiency of human cytochrome P450 CYP3A4 at electrode surfaces through protein engineering. *Electrochem. Commun. 10*, 1744−1747.

(17) Krishnan, S., Wasalathanthri, D., Zhao, L., Schenkman, J. B., and Rusling, J. F. (2011) Efficient bioelectronic actuation of the natural catalytic pathway of human metabolic cytochrome P450s. *J. Am. Chem. Soc. 133*, 1459−1465.

(18) Williams, P. A., Cosme, J., Vinković, D. M., Ward, A., Angove, H. C., Day, P. J., Vonrhein, C., Tickle, I. J., and Jhoti, H. (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science 305*, 683−686.

(19) Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: The biomolecular simulation program. *J. Comput. Chem. 30*, 1545−1614.

(20) MacKerell, A. D., Feig, M., and Brooks, C. L. (2003) Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc. 126*, 698−699.

(21) Foloppe, N., Sagemark, J., Nordstrand, K., Berndt, K. D., and Nilsson, L. (2001) Structure, dynamics and electrostatics of the active site of glutaredoxin 3 from *Escherichia coli*: Comparison with functionally related proteins. *J. Mol. Biol. 310*, 449−470.

(22) Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem. 31*, 671−690.

(23) Vanommeslaeghe, K., and MacKerell, A. D. (2012) Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model. 52*, 3144−3154.

(24) Vanommeslaeghe, K., Raman, E. P., and MacKerell, A. D. (2012) Automation of the charmm general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model. 52*, 3155−3168.

(25) van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free. *J. Comput. Chem. 26*, 1701−1718.

(26) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.

(27) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.

(28) Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 1463−1472.

(29) Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys. 126*, 14101−14107.

(30) Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. 52*, 7182−7190.

(31) Nosé, S., and Klein, M. L. (1983) Constant pressure molecular dynamics for molecular systems. *Mol. Phys. 50*, 1055−1076.

DOI: 10.1021/bi5011656
Biochemistry 2015, 54, 711−721

(32) (2014) *The PyMOL Molecular Graphics System*, version 1.5.0.4; Schrödinger, LLC: Camberley, U.K. http://www.pymol.org/

(33) Seeber, M., Cecchini, M., Rao, F., Settanni, G., and Caflisch, A. (2007) Wordom: A program for efficient analysis of molecular dynamics simulations. *Bioinformatics 23*, 2625−2627.

(34) Vitalis, A., and Caflisch, A. (2012) Efficient construction of mesostate networks from molecular dynamics trajectories. *J. Chem. Theory Comput. 8*, 1108−1120.

(35) Vitalis, A., and Pappu, R. V. (2009) Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem. 5*, 49−76.

(36) Harlow, G. R., and Halpert, J. R. (1997) Alanine-scanning mutagenesis of a putative substrate recognition site in human cytochrome P450 3A4: role of residues 210 and 211 in flavonoid activation and substrate specificity. *J. Biol. Chem. 272*, 5396−5402.

(37) Omura, T., and Sato, R. (1964) Carbon monoxide-binding pigment of liver microsomes: II. Solubilization, purification, and properties. *J. Biol. Chem. 239*, 2379−&.

(38) Tsalkova, T. N., Davydova, N. Y., Halpert, J. R., and Davydov, D. R. (2007) Mechanism of interactions of alpha-naphthoflavone with cytochrome P450 3A4 explored with an engineered enzyme bearing a fluorescent probe. *Biochemistry 46*, 106−119.

(39) Davydov, D. R., Sineva, E. V., Sistla, S., Davydova, N. Y., Frank, D. J., Sligar, S. G., and Halpert, J. R. (2010) Electron transfer in the complex of membrane-bound human cytochrome P450 3A4 with the flavin domain of P450BM-3: The effect of oligomerization of the heme protein and intermittent modulation of the spin equilibrium. *Biochim. Biophys. Acta, Bioenerg. 1797*, 378−390.

(40) Fernando, H., Davydov, D. R., Chin, C. C., and Halpert, J. R. (2007) Role of subunit interactions in p450 oligomers in the loss of homotropic cooperativity in the cytochrome P450 3A4 mutant L211F/D214E/F304W. *Arch. Biochem. Biophys. 460*, 129−140.

(41) French, J. S., and Coon, M. J. (1979) Properties of NADPH-cytochrome *P*-450 reductase purified from rabbit liver microsomes. *Arch. Biochem. Biophys. 195*, 565−577.

(42) Davydov, D. R., Deprez, E., Hui Bon Hoa, G., Knyushko, T. V., Kuznetsova, G. P., Koen, Y. M., and Archakov, A. I. (1995) High-pressure-induced transitions in microsomal cytochrome P450 2B4 in solution: Evidence for conformational inhomogeneity in the oligomers. *Arch. Biochem. Biophys. 320*, 330−344.

(43) Khan, K. K., He, Y. Q., Domanski, T. L., and Halpert, J. R. (2002) Midazolam oxidation by cytochrome P450 3A4 and active-site mutants: An evaluation of multiple binding sites and of the metabolic pathway that leads to enzyme inactivation. *Mol. Pharmacol. 61*, 495−506.

(44) He, Y. A., He, Y. Q., Szklarz, G. D., and Halpert, J. R. (1997) Identification of three key residues in substrate recognition site 5 of human cytochrome P450 3A4 by cassette and site-directed mutagenesis. *Biochemistry 36*, 8831−8839.

(45) Kumar, S., Davydov, D. R., and Halpert, J. R. (2005) Role of cytochrome B_5 in modulating peroxide-supported CY3A4 activity: Evidence for a conformational transition and cytochrome P450 heterogeneity. *Drug Metab. Dispos. 33*, 1131−1136.

(46) Sevrioukova, I. F., and Poulos, T. L. (2013) Understanding the mechanism of cytochrome P450 3A4: Recent advances and remaining problems. *Dalton Trans. 42*, 3116−3126.

(47) Ekroos, M., and Sjögren, T. (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. U.S.A. 103*, 13682−13687.

(48) Sevrioukova, I. F., and Poulos, T. L. (2010) Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc. Natl. Acad. Sci. U.S.A. 107*, 18422−18427.

(49) Park, H., Lee, S., and Suh, J. (2005) Structural and dynamical basis of broad substrate specificity, catalytic mechanism, and inhibition of cytochrome P450 3A4. *J. Am. Chem. Soc. 127*, 13634−13642.

(50) Sevrioukova, I. F., and Poulos, T. L. (2013) Dissecting cytochrome P450 3A4−ligand interactions using ritonavir analogues. *Biochemistry 52*, 4474−4481.

(51) Baas, B. J., Denisov, I. G., and Sligar, S. G. (2004) Homotropic cooperativity of monomeric cytochrome P450 3A4 in a nanoscale native bilayer environment. *Arch. Biochem. Biophys. 430*, 218−228.

(52) Sineva, E. V., Rumfeldt, J. A. O., Halpert, J. R., and Davydov, D. R. (2013) A large-scale allosteric transition in cytochrome P450 3A4 revealed by luminescence resonance energy transfer (LRET). *PLoS One 8*, e83898.

SUPPLEMENTAL

Selection and suggestion of mutants

The selection of mutants was entirely based on the largest clusters from the single CBZ molecular dynamics simulation (Table S1). The prime notion was that the effect of methylation of small molecules, which is done in iterations of drug optimizations to increase potency, can also be achieved by increasing the hydrophobic surface of the protein, since the CBZ had to stay the same. Table S1 shows the centre of mass distances for the CBZ molecule in the largest cluster and the surrounding residues. While side chains featuring a hydrogen bond donor/acceptor were excluded from mutation due to unforeseeable influences on the charge distribution of the active site, residues with a hydrophobic side chain in a certain distance were considered. F residues (already at maximal hydrophobic surface) were not considered either. Only residues close to the 5H-dibenzo [b,f] azepine featuring a side chain which would be closer after mutation could potentially increase the contact of vdW interaction. Considering these criteria, only A370 and I369 were left for mutagenesis. The S119A mutant was suggested to disrupt the only present hydrogen bond and thus the overall stability of the binding mode. I301, A305, and G306 were not considered for mutation since an increase in vdW side chain was assumed to introduce steric clashes and thus obstruct the binding mode. I120 was on the side opposite to the 5H-dibenzo [b,f] azepine and too close to the hydrogen bond with S119 to be considered for mutation.

Table S1: Me**an distance of center of mass (COM) for the respective residues to the COM of carbamazepine in the largest cluster**[*]**.**

residue	mean	residue	mean
Arg105	8.97	Ala297	12.72
Arg106	14.09	Gln298	12.1
Pro107	14.41	Ser299	13.19
Phe108	9.46	Ile300	11.47
Gly109	14.28	Ile301	6.85
Phe110	16.28	Phe302	9.95
Val111	12.12	Ile303	11.43
Gly112	15.41	Phe304	6.01
Phe113	15.19	Ala305	5.45
Met114	10.86	Gly306	9.15
Ile118	9.82	Tyr307	11.96
Ser119	5.83	Glu308	8.28
Ile120	8.5	Thr309	7.85
Ala121	11.79	Leu366	20.52
Leu210	10.36	Phe367	16.9
Leu211	11.05	Pro368	13.16
Arg212	6.38	Ile369	8.01
Phe213	8.81	Ala370	8.74
Asp214	13.47	Met371	12.88
Phe215	13.62	Arg372	14.33
Leu221	20.75	Leu373	11.59
Ser222	22.02	Leu479	21.44
Ile223	18.59	Gly480	16.69
Thr224	19.09	Gly481	13.45
Val225	23.88	Leu482	10.76
Phe226	25.67	Leu483	13.48
Pro227	23.87	Gln484	14.24
Phe228	27.1	Pro485	17.97

[*] Based on the binding mode, the decision was made to investigate the increase of hydrophobic contact without introducing either steric clashes or changing the biochemical property of the residue (e.g. change a Thr to a Phe). This only left A370 and I369 for consideration since other residues were either too close to CBZ or already large enough in terms of vdW size of the side chain.

Derivation of an equation that combines an allowance for partial substrate inhibition with an approximation of cooperative interactions in terms of the Hill equation

The Hill equationwas first introduced in 1910 by A. Hill (Hill, A.V. (1910), J. Physiol. 40 (Suppl): iv–vii) to depict the interactions of hemoglobin with oxygen. Itis based on a simple model where the complex of a protein with ligand is formed via simultaneous binding of *h* molecules of ligand. Although the initial model considered the interactions of a protein oligomer (*h*-mer) with *h* molecules of ligand, the assumption of oligomerization of the protein is not necessary for this formalism to be applicable. The Hill equation in its modern representation may be obtained considering an equilibrium of simultaneous binding of *h* molecules of ligand *S* to protein *E* (either monomer or oligomer, or a subunit of an oligomer) with the formation of complex X:

$$E + hS \xrightleftharpoons{K_D} X$$

The equilibrium constant K_D that defines the steady-state concentration of the complex [*X*] may be represented as:

$$K_D = \frac{([E]_0 - [X]) \cdot [S]^h}{[X]} \tag{1}$$

Resolving this equation for [*X*] for the case where the total concentration of the ligand ($[S]_0$) is much higher than the total concentration of the protein ($[E]_0$), so that $[X] << [S]_0$ and $[S] \approx [S]_0$, we obtain:

$$[X] = \frac{[E]_0 \cdot [S]_0^h}{K_D + [S]_0^h}$$

Now introducing a new parameter S_{50}, defined as $S_{50} = \sqrt[h]{K_D}$ we get the Hill equation in its most common representation:

$$[X] = \frac{[E]_0 \cdot [S]_0^h}{S_{50}^h + [S]_0^h} \tag{2}$$

Let us complement this model with a binding of one more substrate molecule to the complex *X* resulting in the formation of complex X′:

$$E + hS \xrightleftharpoons{K_D} X \xrightleftharpoons{S \searrow K_I} X'$$

Considering the case where $[S]_0 >> [E]_0$ so that $[S] \approx [S]_0$, we obtain the following relationship for the dissociation constant K_D of the complex *X*:

$$K_D = \frac{([E]_0 - [X] - [X']) \cdot [S]^h}{[X]} \tag{3}$$

The dissociation constant of the (inhibitory) complex *X′* (K_I) may be defined as:

$$K_I = \frac{[X] \cdot [S]}{[X']} \tag{4}$$

Resolving the system of equations (3) and (4) for [*X*] and [*X′*] and introducing the parameter S_{50}, which is defined as $S_{50} = \sqrt[h]{K_D}$, we obtain the following equations:

$$[X] = \frac{K_I \cdot [E]_0 \cdot [S]^h}{[S]^h \cdot (K_I + [S]) + S_{50}^h \cdot K_I} \tag{5}$$

and

$$[X'] = \frac{[E]_0 \cdot [S]^{h+1}}{[S]^h \cdot (K_I + [S]) + S_{50}^h \cdot K_I} \tag{6}$$

Let's now consider the case when the protein E is an enzyme, ligand S is a substrate, the complex *X* is the enzyme-substrate complex with the maximal rate of catalytic turnover, and the complex *X′* is an inhibitory complex, where the rate of catalysis is decreased due to the binding of an additional (inhibitory) molecule of substrate. We define the parameter α as a ratio of the turnover numbers (apparent catalytic constants) characteristic to the complexes *X* (k_{cat}) and *X′* (k'_{cat}):

$$\alpha = \frac{k'_{cat}}{k_{cat}}$$

We may represent the overall rate of catalysis (*V*) as a function of [*X*] and [*X′*]:

$$V = k_{cat} \cdot [X] + k'_{cat} \cdot [X'] = V_{\max} \cdot ([X] + \alpha \cdot [X']) \tag{7}$$

Combining equations (5), (6) and (7) we obtain:

$$V_S = \frac{k_{cat} \cdot [E]_0 \cdot [S]^h \cdot (\alpha \cdot [S] + K_I)}{[S]^h \cdot (K_I + [S]) + S_{50}^h \cdot K_I} \tag{8}$$

This equation is identical to the Equation 2 in the manuscript. Considering the case when cooperativity of the interactions is eliminated (h = 1) and replacing S_{50} with K_M we can obtain Equation 1 of the manuscript.

Similar to the regular Hill equation, equation (8) is based on a presumption that the binding of several ligand molecules to the protein takes place simultaneously. Due to this outermost simplification, the parameter h (the Hill coefficient) cannot be straightforwardly considered as a number of binding sites involved in the formation of complex X. However, similar to the case of the regular Hill equation, it may be shown that, for a mechanism with a positive cooperativity ($h > 1$), this parameter is less or equal to the actual number of binding sites involved in the interactions (see: A. Cornish-Bowden, Fundamental of Enzyme Kinetics, Butterworths, 1979, Chapter 7).

Distances of the CBZ C10,11 COM to the heme iron in MD Simulations with 2-4 CBZ molecules

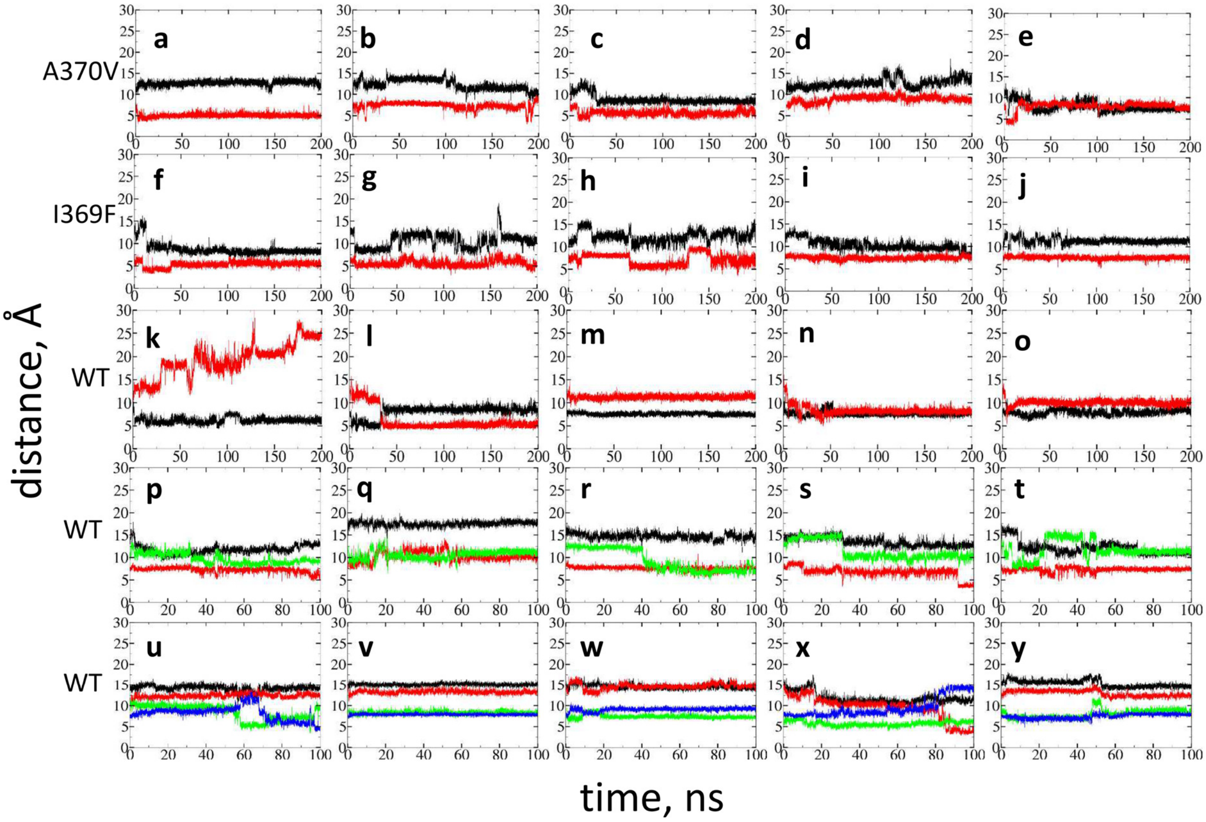


Figure S1: Time series of the distance between the center of the C10-C11 bond in CBZ and the heme iron. Each color corresponds to a single CBZ. Five 200 ns simulations were performed with two CBZs with the A370V mutant (a-e), the I369F mutant (f-j), and the wild-type CYP3A4 (k-o); five 100-ns simulations were performed with the wild-type with three (p-t) and four (u-y) CBZ molecules. All figures were generated with xmgrace. No full dissociation event is observed in the MD runs. The largest separation of CBZ from the heme is observed in panel (k) and a snapshot from the corresponding MD run is shown in Fig. 7 in the main section.

Distances between the side-chains of the residues 369 and Phe304

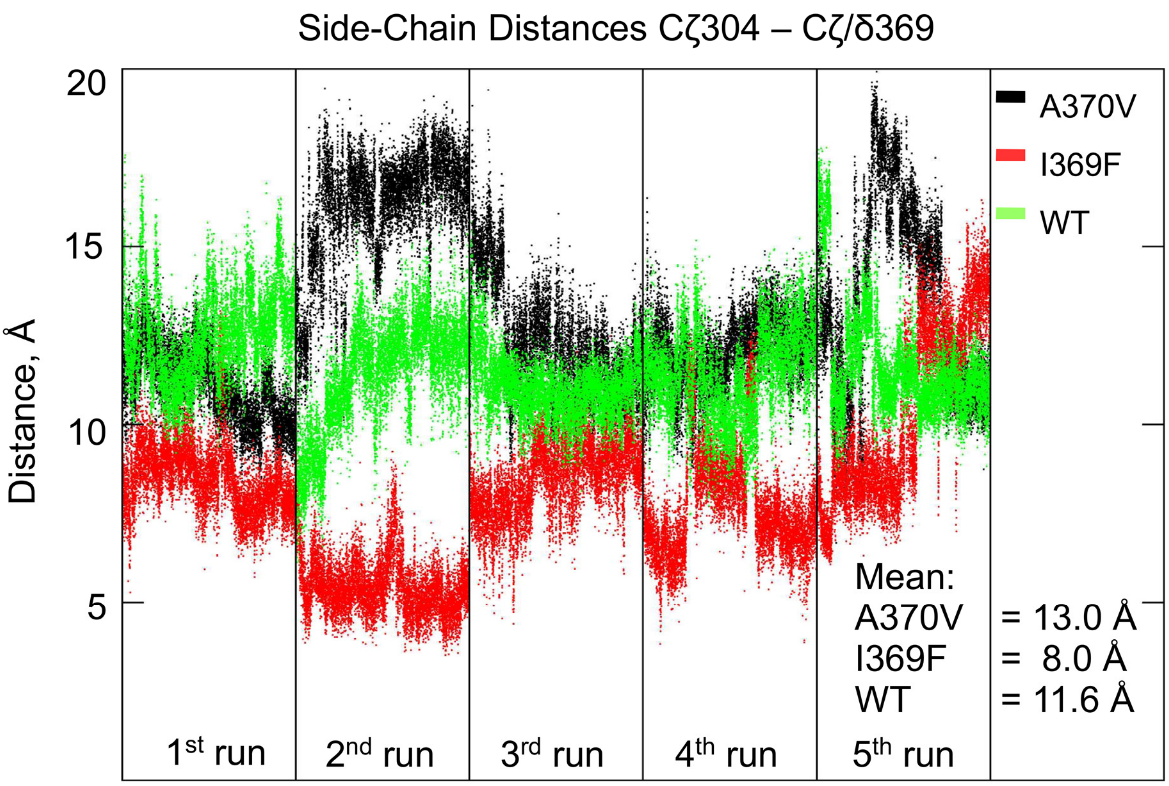


Figure S2: Time-series of distances between the side chains of residue 369 and Phe304 of the five simulations of a complex of two CBZ molecules with CYP3A4 wild-type (light green) and the mutants A370V (black) and I369F (red). The data shown here are based on the identical simulations as the evaluations of Fig. S1 (a-o).

Chapter 4

Discovery of a SMARCA4 bromodomain inhibitor by fragment-based high-throughput docking

Tim Knehans, Amedeo Caflisch

Manuscript in preparation

Discovery of a SMARCA4 bromodomain inhibitor by fragment-based high-throughput docking

Tim Knehans, Amedeo Caflisch*

University of Zurich, Department of Biochemistry, Winterthurerstrasse 190, 8057 Zurich, Switzerland

*** Corresponding author:**
Amedeo Caflisch,

caflisch@bioc.uzh.ch

Introduction

With the increased recognition that proteins involved in epigenetic regulation are a promising target for drug development against developmental diseases such as cancer, efforts to find small molecule inhibitors against epigenetic targets have been increased significantly[1–3]. A prime focus of research in this area has been the role and subsequent interference of bromodomains (BRD)[4–7]. Bromodomains have been identified as an interface for histone recognition in larger protein complexes[8]. SMARCA4/BRG1 is one of two mutually exclusive helicases of the SWI/SNF chromatin remodeling complex. In addition to SWI/SNF, SMARCA4 was also identified as being part of WINAC, NUMAC, NCoR and mSin3A/HDAC complexes[9]. The helicase domain represents the central catalytic domain of these complexes and is speculated to be involved in the transcription and repression of a multitude of genes, in a contextual manner. As part of SWI/SNF, SMARCA4 is crucial in early embryonic development and spermatogenesis. Mutations inactivating SMARCA4 which mostly occur in the helicase domain have been identified in several ovarian, breast, lung, pancreas and prostate cancer lines well as Coffin-Syris syndrome[10–16]. Contrary to the detrimental effects of loss of SMARCA4 activity in mutated genes, inhibition of functional SMARCA4 becomes significant as a tumor suppressor mechanism in the context of SWI/SNF mediated cancers[17]. Wang et al. (2009) have shown that upon loss of the SNF5 subunit of SWI/SNF, oncogenesis is dependent on the retained activity of SMARCA4[18]. Further, investigations into Wnt-dependent tumorourigenesis indicated that loss of SMARCA4 functionality slows aberrant Wnt-signalling in mice[19]. In addition to the helicase, SMARCA4 also contains a class VIII bromodomain which acts as the recognition module for acetylated lysines (Kac) which loses the capacity for binding upon mutation of the conserved ASN1540[20].

As of today, there is a total of three structures available for the SMARCA4 bromodomain, two crystal structures (PDB-ID: 3UVD, 2GRC) and one NMR ensemble (PDB-

ID: 2H60). Out of these structures only 3UVD contains a fragment-sized ligand namely 1-methylpyrrolidin-2-one (PDB ligand identifier MB3). The complex features a single hydrogen bond between the MB3 carbonyl oxygen and the side chain amide of Asn1540 at a distance of 2.5 Å and an H-bond angle of 155°. Molecular dynamics studies conducted for series of bromodomains including SMARCA4 assessing the $\chi2$ and $\chi1$ dihedral dynamics of its conserved asparagine (Asn1540) and the N-adjacent Phe1539 respectively showed several rotational events for both residues within a single 0.5 µs simulation. The study also showed a change of binding site accessible molecular surface (SASA) upon the change of Phe1539 $\chi1$. Small molecules with low µM K_i values for SMARCA4 have been primarily identified only in the context of a bromoscan or selectivity panels for a series of molecules designed specifically for BRD4(1) or BRPF1 (Figure 1a-c)[21,22]. Additionally, thermal shift data for a series of 2-thiazolidinones derivatives, RVX-208 derivatives as well as dual kinase-bromodomain inhibitors showed ΔT values below 1°C at 10 µM and 50 µmol of compound respectively[23–25]. However, to this day PFI3 is the only dedicated chemical probe reported for SMARCA4 , albeit not published in the scientific literature (Figure 1d)[26]. The Structural Genomics Consortium (SGC) reported estimations of PFI3 dissociation constants (K_D) via isothermal titration calorimetry (ITC) at 89 nm for SMARCA4. Thermal shift experiments are reported as showing a 5.1° shift for SMARCA4 in the presence of 10 µM PFI3.

Here we report the identification of a ligand of the SMARCA4 bromodomain by an *in silico* fragment-based approach[27]. To our knowledge SMARCA4 has not yet been directly targeted by any virtual screening campaign.

Methods

Selection and preparation of SMARCA4 structure for docking

At the start of this project, three structures of the SMARCA4 bromodomain were available at the RCSB protein databank (PDB-ID: 3UVD, 2GRC, 2H60)[28]. The only structure containing a small-molecule ligand (3UVD) was chosen for this study[29]. Except for four structural waters in the Kac binding site (numbered 2, 4, 19, and 58 in 3UVD), all water molecules and 1-methylpyrrolidin-2-one were removed. The amino group of the N-terminal Leu1458 was considered as positively charged and the carboxyl group of the C-terminal Lys1563 as negatively charged. Coordinates of hydrogen atoms and missing atoms of the Lys1460, Lys1461, Lys1473, Lys1492, Glu1493, and Leu1541 side chains were generated by the *psfgen* module of VMD[30]. The CHARMM36 and TIP3P parameters were used for SMARCA4 and the four water molecules, respectively[30,31]. The coordinates of the hydrogen atoms were minimized in CHARMM first by 2,000 steps of steepest descent followed by the conjugate gradient algorithm (20,000 steps) using as convergence criterion a gradient of the energy of 0.01 kcal/(mol Å)[33–35].

Library generation:

The ZINC 12 all-purchasable library (version 12 of 01 2012, containing 22,724,825 molecules) was downloaded from http://zinc.docking.org/ with protonation states at neutral pH[36]. The molecules were automatically decomposed by the program DAIM which cuts at rotatable bonds[37]. The program MATCH was used to generate CHARMM General Force Field (CGenFF v.2b6) parameters for all fragments including 1-methylpyrrolidin-2-one from the 3UVD structure (which is a positive control for docking)[38,39]. As further controls, a library of known bromodomain inhibitors was assembled by merging the ChEpiMod database with remaining published inhibitors and patented structure not available in the ChEpiMod at the time[40]. ChEpiMod contained a multitude of molecules tested for the BAZ2B bromodomain labeled as inactive/inconclusive at a given concentration, these molecules were removed from the database.

Fragment docking and fragment selection:

All rigid fragments (no rotatable bonds or a single hydroxyl group) were docked by SEED (v. 3.3.6.) with MATCH-generated CGenFF atom types. The dielectric constant of the solute was set to 2.0, and default geometric parameters were used[38,39,41]. The binding site used for docking included Val1484, Phe1485, Ile1486, Gln1487, Leu1488, Pro1489, Tyr1497, Ala1536, Phe1539, N1540, and Ile1546 as well as waters 2, 4, 19 and 58. All fragments were docked on Xeon 5560 processors clocked at 2.8 GHz each. The co-crystallized 1-methylpyrrolidin-2-one was evaluated for its total SEED energy using the pose as in the X-ray structure, i.e., without prior minimization of the complex. The calculated total SEED energy of 1-methylpyrrolidin-2-one was then set as the lower boundary for fragment filtering at -8.16 kcal/mol.

All fragments with a lower total SEED energy were then evaluated for hydrogen bonds with the SMARCA4 binding site residues as defined for SEED docking. The criteria for hydrogen bonds were a distance between donor and acceptor heavy atoms lower than 3 Å and the hydrogen bond angle (donor-H ••• acceptor) deviating by a maximum of 30 degrees from the optimal angle of 180°.

Substructure search:

All fragments were looked up in larger molecules via RDKit module *HasSubstructMatch(MolFromSmarts([SMARTS]))*[42]. SMARTS patterns with explicit hydrogens involved in hydrogen bonding were searched in the ZINC all-purchasable library. The resulting molecules were then evaluated for their purchasability. Furthermore, all remaining molecules were filtered for containing reactive, unwanted chemical moieties (Supplementary 1) derived from lists provided by Hann et al. 1999 and Brenk et al. 2007[43,44].

Tethered docking:

The minimized SMARCA4 structure from the fragment stage, including the four water molecules, were split into individual pdb files and served as the receptor and structural waters. To keep the original fragment in place, the anchor fragments where tethered via *sdtether* supplied with rDock[45,46]. The SMARTS pattern required for the previous substructure search was supplied to *sdtether* along the mol2 (TRIPOS atom types) files of the respective molecules and the fragment mol2 file containing its docked coordinates[47]. Tethered docking was performed on Xeon 5560 clocked at 2.8 GHz each. In the *LIGAND* section of the rDock parameter file, translation and rotation (*TRANS_MODE*, *ROT_MODE*) were kept tethered with a maximal translation and rotation of 1 Å and 30° respectively. The dihedrals were allowed to freely rotate (*DIHEDRAL MODE FREE*) to account for eventual accommodation of hydroxyl groups. The structured waters in the *SOLVENT* section were also treated as *TETHERED* in both rotation and translation with maximal values similar as for the *LIGAND* section. The *OCCUPANCY* was set to 1 Å. The site mapper was set to *RbtSphereSiteMapper* with a center at (6.049, 14.390, 44.171), the *GRIDSTEP* was set to 0.5 Å with a radius of 11.0 Å. The small sphere for site mapping was set to a radius of 1 Å while the large sphere setting (*LARGE_SPHERE*) was set to a radius of 7 Å. A single cavity (*MAX_CAVITIES* 1) was chosen. After docking, all conformations having a positive energy (<*SCORE*>) were discarded from further evaluation.

CHARMM minimization and calculation of electrostatic solvation:

Following docking in rDock, non-polar hydrogens were added in openbabel (version 2.3.2)[48]. All poses obtained from rDock for each molecule were evaluated for the number of hydrogen bonds it formed with SMARCA4 residues. Subsequently, the 36 molecules featuring four hydrogen bonds were parameterized for CGenFF through Paramchem[49,50]. Due to upload restrictions of the Paramchem web service, molecules featuring conformations with two and three hydrogen bonds (a total of 4198 and 159,220, respectively) were parameterized for CGenFF via MATCH[38]. The generated rtf(topology), prm(parameter), and str(stream) files were supplied to *psfgen* module from VMD in order to generate psf files and pdb files for the respective molecules, containing the docked coordinates and individual parameters necessary for processing in CHARMM. The docked molecules were then minimized in the receptor by CHARMM using first the steepest descent algorithm (500 steps) followed by the conjugate gradient algorithm (10,000 steps) with a convergence criterion of 0.01 kcal/(mol Å). After minimization the intermolecular van der Waals energy (ΔE_{vdw}) and Coulombic energy ($\Delta E_{Coulomb}$) were evaluated. The electrostatic solvation energies (for the complex, ligand, and protein individually) were calculated by the finite-difference Poisson module in CHARMM using dielectric constants of 2.0 and 78.2 for the solute and solvent, respectively[51]. The electrostatic contribution to solvation energy of the binding process (ΔE_{solv}) was then approximated as the sum of the electrostatic solvation energies of the complex, receptor, and ligand. The electrostatics component of the free energy of binding (ΔG_{elec}) was calculated as the sum of $\Delta E_{Coulomb}$ and ΔE_{solv}. Additionally, ΔElect was calculated as a measure of how much the ligand-bound state is electrostatically favorable compared to the ligand in solvent. ΔElect is computed as the difference between intermolecular electrostatic interaction of the ligand in solvent ($E_{inter,solv}$) calculated by the finite-difference Poisson module in CHARMM using dielectric constants of 78.2 and the electrostatic free energy of solvation of the ligand ($\Delta E_{solv,lig}$). Finally, the total interaction energy E_{total} was calculated as the sum of ΔE_{vdw} and ΔG_{elec}. Following, all molecules with a positive E_{total} were discarded. All 15 molecules featuring 4 and 3 hydrogen bonds before energy minimization where visually inspected after energy minimization. Molecules featuring 2 H-bonds had all molecules removed which had a positive ΔElect and were then ranked according to E_{total}. Out of these 195 remaining molecules (180 with 2 H-bonds, 7 with 3 H-bonds, 8 with 4 H-bonds), a set of 32 molecules for compound purchase was assembled.

Alpha Screen Assay:

Prior to measuring single dose and dose-response values, all compounds were dissolved/diluted in DMSO to a concentration of 50 µM (with the exception of ZINC22642290 which was diluted to 25 µM). Both, single-dose and dose-response measurements determining the half maximal inhibitory concentration (IC_{50}) for each compound were performed by Reaction Biology Corp. deploying the ALPHA Screen binding assay. In this assay donor beads coated with streptavidin were incubated with biotinylated histone H3 (residues 1-30) containing Kac (K14/18/23/27Ac-GG). Fluorescence at 520-620 nm resulting from the close proximity of Kac with SMARCA4 after excitation at wavelength of 680 nm was then measured in the EnVision^{TM} multilabel plate reader for solutions containing the ligand and without. IC_{50} measurements for PFI3 (control) and ZINC04181101 were tested in 10-dose mode with 3-fold serial dilution starting at 50 µM.

Results

Structure selection and binding site evaluation:

For SMARCA4 there are currently three available 3D-structures out of which two structures were obtained by means of X-ray crystallography(3UVD, 2GRC) and one by NMR spectroscopy(2H60). Evaluation of the RMSD for all heavy atoms between 3UVD and 2GRC is 0.27 Å while between 3UVD and 2H60 the RMSD is 1.85 Å. A different orientation of the aZ'-aZ loop in 2H60 is the conformational feature which contributes most to the higher RMSD between 3UVD and 2H60 while the binding site is highly similar in conformation (Supplementary 1). The decision for utilizing 3UVD for the virtual screening campaign was finally based upon this being a co-crystal structures with bound 1-methylpyrrolidin-2-one which served as the reference compound for the fragment stage (Figure 2b). Estimating the interaction energy in SEED for the crystal conformation of 1-methylpyrrolidin-2-one resulted in a total energy of -8.16 kcal/mol which was the cut-off value do discard any docked fragment with a higher SEED total energy. Compared to other bromodomains, SMARCA4 provides mainly hydrophobic residues in its binding site (Figure 2a, c). Other than Tyr1497 and Asn1540 which are conserved among the majority of bromodomains only Gln1487 features side chain hydrogen bond donors within 14 Å of 1-methylpyrrolidin-2-one. However, analysis of the flexibility of the simulations previously run on SMARCA4 by Steiner et al. 2013 showed that the Gln1487 sidechain is freely moving and thus does not provide any stable or meta-stable conformation which could be selected to guide structure based design (Supplementary 2). All remaining residues in the Kac binding site are Val148, Phe1485, Ile1486, Leu1488, Pro1489, Leu1494, Val1505, Ala1532, Phe1539 and Ile1546 (Figure 2c). Hydrogen bond acceptors are contributed by the backbone carbonyl atoms of Val1484, Gln1487, and Val1505. The backbone conformation of 3UVD does not provide for any hydrogen bond donors which could be exploited.

Fragment generation and fragment docking:

The ZINC library still constitutes the largest available chemical available for virtual screening campaigns. At the beginning of the *in silico* campaign (10 2014), version 12 of the ZINC database contained 22,724,825 molecules which were fragmented at every heavy atom – heavy atom bond into 204,754 unique fragments. As SEED is not able to rotate dihedral angles in molecules, at first only molecules featuring no rotatable bonds were selected for fragment docking. However, as hydroxyl groups are both rotatable and provide a very potent H-bond donor-acceptor-pairs, molecules containing hydroxyl groups were kept. This yielded a fragment library of 41,623 fragments eligible for docking in SEED. Since the CHARMM General Force Field (CGenFF) currently represents the most versatile small molecule parameters, all fragments were parameterized for CGenFF via MATCH which resulted in a net loss of 7,548 molecules which could not be parameterized due to the unavailability of parameters as well as substitute parameters, bond increment rules for calculation partial charges and inadequate fragments derived from parent ZINC molecules. After docking a total of 37,075 fragments into 3UVD, 32,746 molecules passed the cut-off set by the reference value of -8.16 kcal/mol of 1-methylpyrrolidin-2-one. Hydrogen bonds provide a significant interface for the interaction of small molecules with their respective protein target. As the strength of a hydrogen bond is defined by distance between the donor/acceptor heavy atoms and the angle between the donor pair and acceptor atom, more rigid distance and angle criteria of 2.5 - 3.0 Å and 180° (±30°) respectively were chosen. All SMARCA4-fragment-complexes were evaluated for at least a single H-bond, resulting in 1,057 fragments.

Substructure Search:

In accord with the ALTA procedure, a substructure search was performed via RDKit to identify potent molecules containing the promising fragments identified in the fragment stage. To keep the moieties involving H-bonds intact and simultaneously allow for more combinatorial possibilities, all polar hydrogens which provided the interface for a H-bond were kept while all other hydrogens were deleted. This yielded a total of 1,165,191molecules representing a chemical space reduction by a factor of 10^1 from the size of the original size of the ZINC set. Unsurprisingly, phenol was the most common fragment identified in 458,510 molecules. To further reduce the chemical space, chemically unfavorable moieties (Supplementary 3) were searched for by substructure search and molecules containing them were removed from the set. As solubility is a prevalent issue in all stages of drug development not only the “Hit phase”, a clogP < 5 was enforced while all other Lipinski rules were disregarded to not further restrict the chemical space. Finally, all molecules were checked for actual availability and vendor information on the ZINC database leaving a total of 921,779 parent molecules ready for docking, energy evaluation, and subsequent ordering. All molecules unavailable for purchase despite displaying promising *in silico* properties otherwise were set aside for later virtual screening campaigns.

Docking, Scoring, and Selection:

The prime advantage of the ALTA procedure and fragment-based design generally, lies in the extension of potent fragments into parent molecules of higher potency while keeping the fragment in its original conformation. A side effect is the reduction of chemical space by removing all molecules which do not contain the wanted fragments. To achieve this goal we incorporated the capabilities or rDock for tethered docking into the design process. Tethered docking allowed for the initially docked fragment to maintain its relative position from the fragment stage, while the rest of the molecule could move freely. Subsequent to docking, all fragments were evaluated via the rDock score and molecules having a positive score indicating unfavorable poses were discarded while molecules passing the threshold were then evaluated for the amount of hydrogen bonds they formed with SMARCA4. The side chain carbonyl oxygen and amide hydrogens of Asn1540 provide a H-bond donor/acceptor pair in a very short distance and interaction with both of them was preferred thus, all molecule containing only a single H-bond were not considered for the subsequent steps. As the residues considered for docking in the Kac binding site of SMARCA4 do not contain a single charged residue, all molecules with a non-zero net charge were removed. Deploying interaction energy calculations with continuum electrostatics in CHARMM for 112,950 molecules resulted in 759 molecules featuring a negative E_{total}. This number was further reduced by requiring a negative ΔElect for molecules with 2 H-bonds prior to minimization which lead to a final set of 180 molecules which were visually inspected. Only 7 (3 H-bonds) and 8(4 H-bonds) molecules had a negative E_{total} after CHARMM minimization and energy evaluation. It is worthwhile to point out that many of these molecules had a positive ΔElect due to the amount of non-carbon atoms, resulting in a significant amount of hydrogen bond donors/acceptors per molecule. Finally, a selection of 32 molecules were then tested (Table 1) in a single dose ALPHA screen.

In vitro ALPHA screen and putative binding mode:

A total of 32 molecules were purchased and then selected for single dose testing via ALPHA screen. At 50 μM, controls Bromosporine and PFI showed 43.6 % and 14.7 % relative signal (Table 2) while ZINC04181101 showed 14.4 %. Reduction of signal below 78 % was not detected for any other compounds tested. A dose response series measured for PFI3 resulted in an IC_{50} value of 271 nM while ZINC04181101 had an IC_{50} value of 20 μM. ZINC04181101 represents a scaffold connecting two ring systems, a phenol and trans-glycoluril, by a N-methleneaminoacetamide linker. Assessment of the absorption spectrum of this compound showed no indication of interference at the wavelengths the ALPHA screen is performed (Supplementary 5). ZINC04181101 is one of 4 molecules which featured 4 hydrogen bonds with SMARCA4 after docking with rDock. Out of which one was removed after energy minimization via CHARMM. Despite this compound having a formal charge of 0, ΔElect was 12.4 due to the number of hydrogen bond donors and acceptors (Table 1). The minimized position shows that ZINC04181101 forms three hydrogen bonds (Figure 5). Two of these are contributed through the trans-glycoluril moiety with the sidechain Asn1540 while the trans-glycoluril group forms another hydrogen bond with the carbonyl backbone oxygen of Gln1487. The third hydrogen bond between ZINC04181101 is formed via the N-methleneaminoacetamide group with Asn1540. Despite an additional hydrogen bond donor/acceptor pair provided by the phenol group, no other H-bond is formed.

The ligand efficiency (LE) calculated according to (1) of ZINC04181101 is 0.28 while the ligand lipophilic efficiency (LLE) calculated via (2) is 1.82.

$$LE = \frac{1.37}{N}\, pIC_{50} \qquad (1)$$

Where N is the number of heavy atoms and pIC_{50} is the negative decadic logarithm in M.

$$LLE = pIC_{50} - cLogP \qquad (2)$$

pIC_{50} as above while the cLogP was estimated as 2.87 according to the method of Viswanadhan et al. 1989 implemented in the Marvinsketch (version 14.11.24.0, Chemaxon http://www.chemaxon.com) node for Knime[53,54]. The authors like to point out that the calculated LogP according to Wildman & Crippen 1999 as implemented in RDKit is -1.16 and the reported xLogP on the ZINC webserver for entry ZINC04181101 is 0.11[55]. Therefor the LLE can change significantly based on the method of LogP calculation.

Conclusions

We have presented the *in silico* identification and *in vitro* validation of a small molecule displaying low µM activity for the SMARCA4 bromodomain. The *in silico* screening has involved fragment docking by the program SEED followed by tethered docking (using rDock) of the subset of molecules that contained one or more of the top fragments. The reduction of a multi-million library of compounds to 32 molecules for *in vitro* assessment was achieved by applying as threshold the calculated binding energy of a structurally resolved fragment as well as hydrogen bond enumeration, van der Waals interaction energy, and the electrostatic contribution to the free energy of binding using the continuum dielectric approximation. The trans-glycoluril moiety provides two out of three hydrogen bonds with Gln1487 and Asn1540, and might serve as an anchor point for future derivatization. It would be worthwhile to investigate if substituting the trans-glycoluril into a cis-glycoluril results similar activity. As the phenol group does not seem to provide significant interaction, it is there were derivatives should be first considered. Removal of the phenol group would increase the ligand efficiency due reduction of the molecules by 6 aromatic carbons. Despite the projected low druggability of SMARCA4, PFI3, the 2-thiazolidinone derivatives by Zhao et al. 2015 and the here presented ZINC04181101 represent three distinct scaffolds which represent a diverse series of scaffolds worthwhile for optimization in order to provide drug candidates to tackle SMARCA4 and SWI/SNF mediated diseases[56–58].

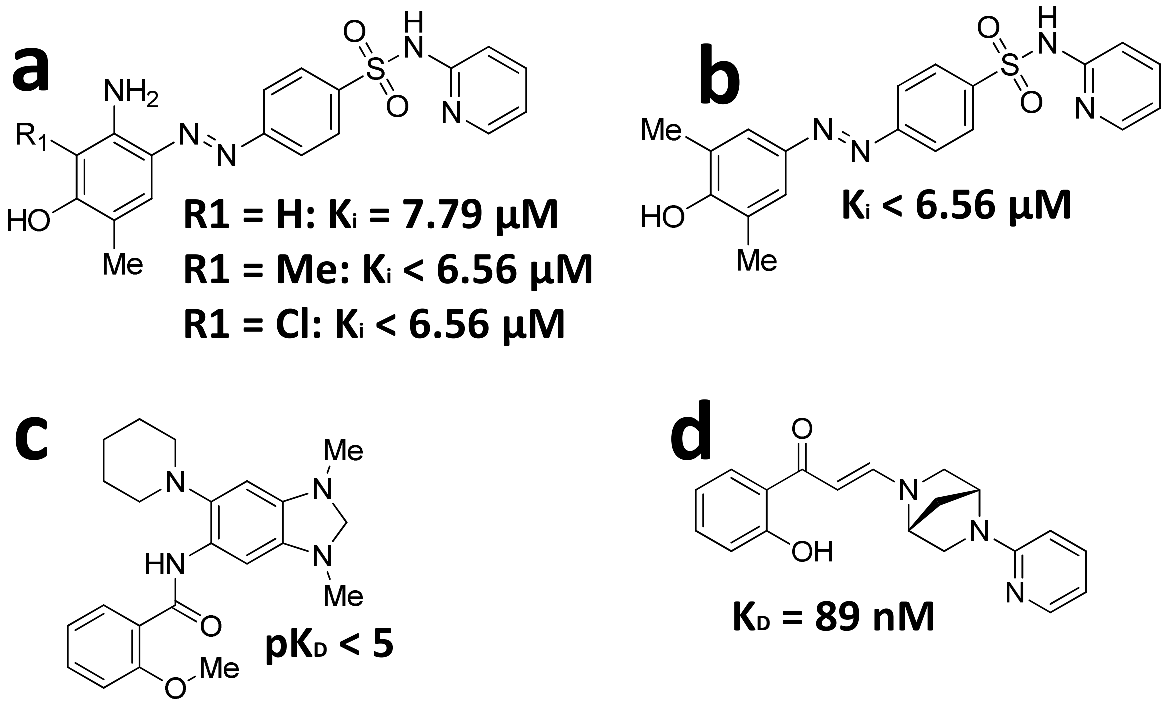
a
R1 = H: K_i = 7.79 µM
R1 = Me: K_i < 6.56 µM
R1 = Cl: K_i < 6.56 µM
b
K_i < 6.56 µM
c
pK_D < 5
d
K_D = 89 nM

Figure 1

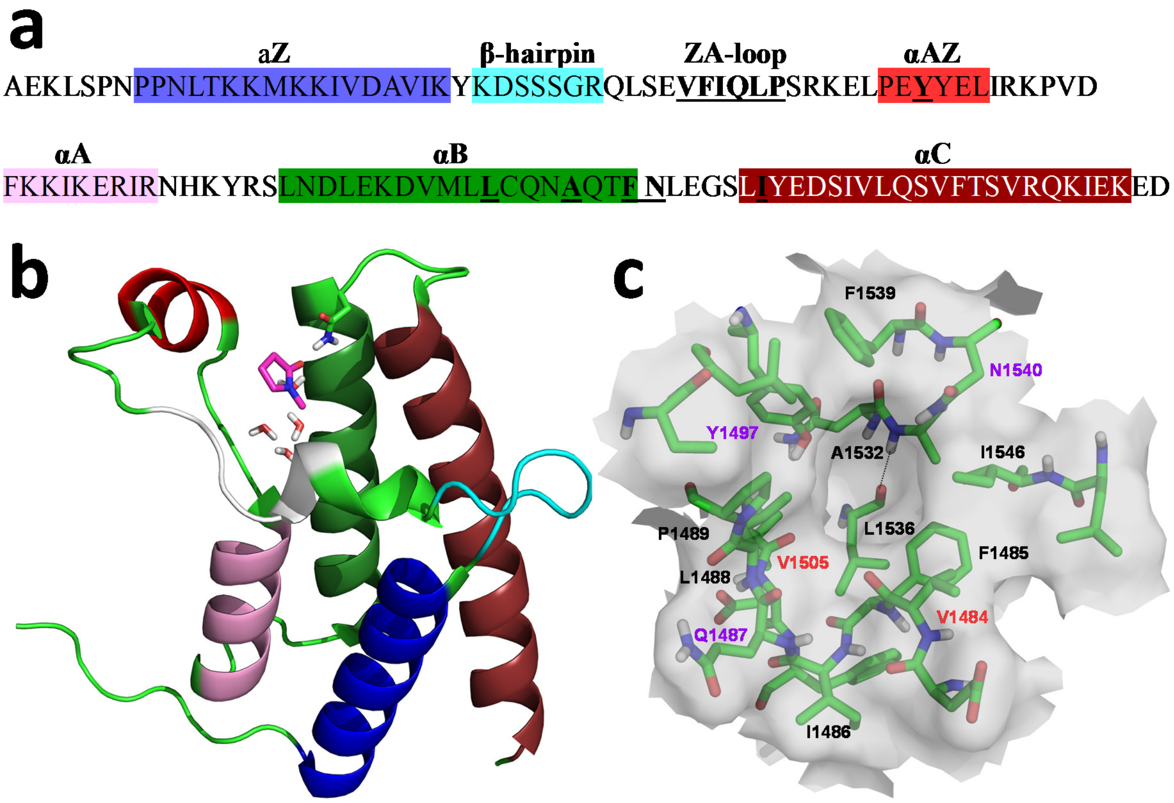

a
aZ
β-hairpin
ZA-loop
αAZ
AEKLSPNPPNLTKKMKKIVDAVIKYKDSSSGRQLSEVFIQLPSRKELPEYYELIRKPVD
αA
αB
αC
FKKIKERIRNHKYRSLNDLEKDVMLLCQNAQTF NLEGSLIYEDSIVLQSVFTSVRQKIEKED
b
c
F1539
N1540
Y1497
A1532
I1546
L1536
P1489
F1485
V1505
L1488
V1484
Q1487
I1486

Figure 2

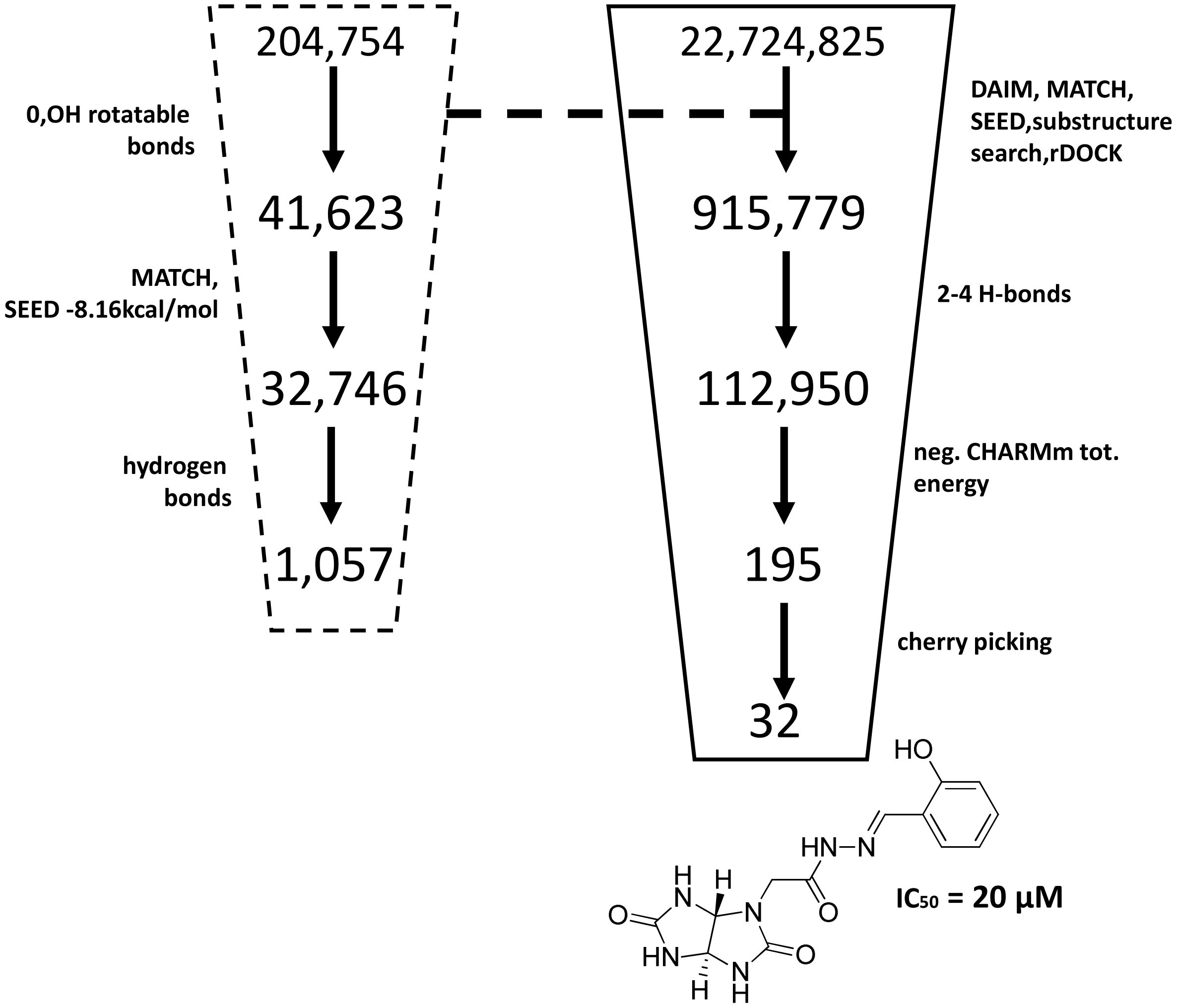

204,754
0,OH rotatable bonds
41,623
MATCH, SEED -8.16kcal/mol
32,746
hydrogen bonds
1,057
22,724,825
DAIM, MATCH, SEED,substructure search,rDOCK
915,779
2-4 H-bonds
112,950
neg. CHARMm tot. energy
195
cherry picking
32
HO
HN–N
H
N
H
O
N
O
HN
H
N
H
O
IC50 = 20 µM

Figure 3

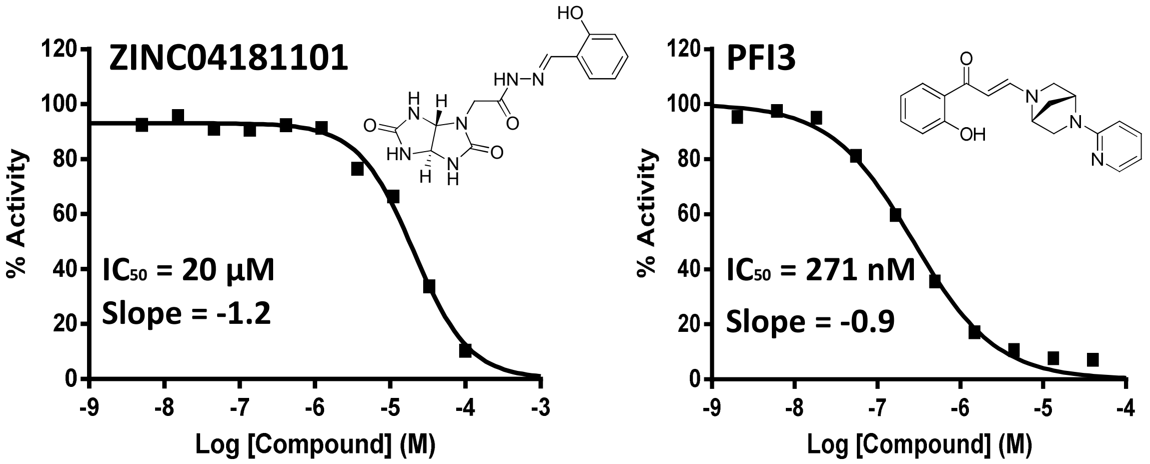

ZINC04181101
IC₅₀ = 20 µM
Slope = -1.2
% Activity
0
20
40
60
80
100
120
-9
-8
-7
-6
-5
-4
-3
Log [Compound] (M)
HO
HN–N
O
N
H
O
HN
N
H
O
PFI3
IC₅₀ = 271 nM
Slope = -0.9
% Activity
0
20
40
60
80
100
120
-9
-8
-7
-6
-5
-4
Log [Compound] (M)
O
OH
N
N
N

Figure 4

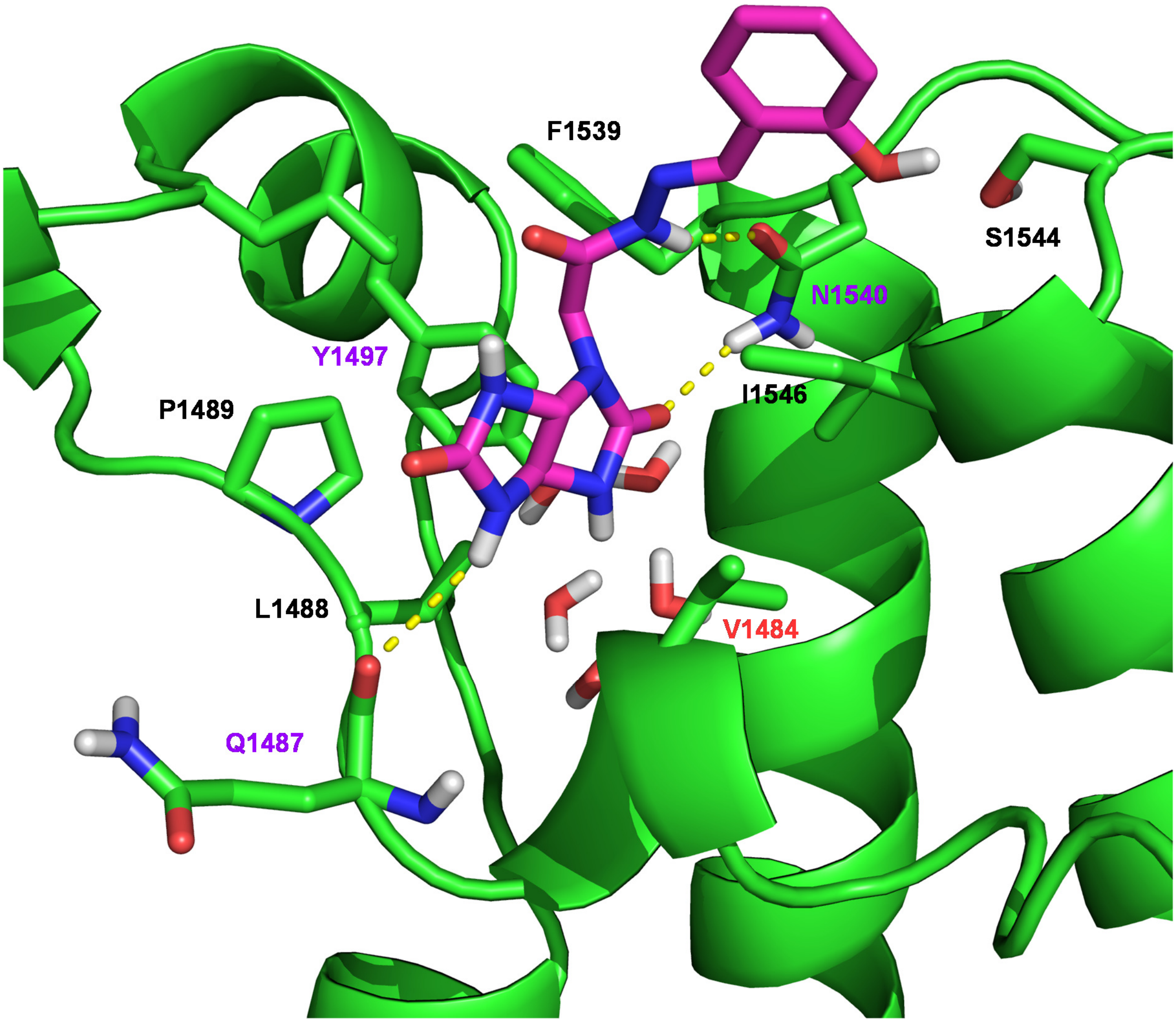


Figure 5

ZINC ID	H-bonds	ΔE_{vdw}	$\Delta E_{coulomb}$	ΔE_{solv}	ΔG_{elec}	$\Delta E_{solv,lig}$	$E_{inter,solv}$	ΔElect	E_{total}
ZINC39404578	2	-35.5	-12.8	17.0	4.2	-9.6	-11.3	-1.8	-31.3
ZINC66699837	2	-32.0	-14.4	22.0	7.6	-10.8	-11.5	-0.6	-24.4
ZINC12863354	2	-27.2	-3.5	7.6	4.1	-7.0	-7.6	-0.6	-23.2
ZINC42735712	2	-25.7	-5.3	8.8	3.5	-7.8	-8.4	-0.7	-22.2
ZINC84678055	2	-25.8	-3.6	7.9	4.2	-5.7	-6.0	-0.3	-21.5
ZINC71920814	2	-26.0	-9.2	14.2	5.0	-9.2	-9.6	-0.4	-21.0
ZINC12858178	2	-23.6	-4.9	7.7	2.8	-5.7	-6.5	-0.8	-20.8
ZINC22642290	2	-25.0	-10.4	14.6	4.2	-8.4	-8.8	-0.4	-20.7
ZINC04940185	2	-23.5	-5.7	9.4	3.7	-6.1	-6.3	-0.3	-19.8
ZINC12901687	2	-26.2	-8.7	15.5	6.8	-6.4	-6.6	-0.2	-19.4
ZINC05417915	2	-25.4	-8.4	14.5	6.1	-7.5	-8.8	-1.3	-19.3
ZINC14121148	2	-23.6	-6.3	10.8	4.5	-5.9	-6.7	-0.8	-19.1
ZINC95356471	2	-19.9	-4.6	5.5	0.8	-7.1	-7.2	-0.1	-19.0
ZINC95348176	2	-21.3	-3.3	5.7	2.4	-8.2	-8.3	-0.1	-18.9
ZINC32855766	2	-22.6	-2.7	6.8	4.1	-5.4	-5.5	-0.1	-18.5
ZINC00127673	2	-20.8	-2.7	5.1	2.4	-5.2	-5.8	-0.5	-18.4
ZINC00076820	2	-21.1	-4.6	7.5	2.9	-5.1	-6.2	-1.1	-18.2
ZINC04206597	2	-20.5	-2.8	5.2	2.4	-5.3	-5.8	-0.5	-18.1
ZINC12804112	2	-19.8	-3.0	5.2	2.2	-5.7	-5.8	0.0	-17.6
ZINC41121701	2	-19.9	-7.0	9.5	2.5	-5.5	-5.9	-0.4	-17.5
ZINC33207842	2	-21.1	-10.3	14.1	3.8	-8.1	-8.2	-0.1	-17.3
ZINC91264242	2	-22.1	-6.4	11.4	5.0	-10.4	-11.2	-0.8	-17.2
ZINC16946449	2	-18.7	-6.8	8.4	1.6	-5.9	-8.3	-2.4	-17.1
ZINC04206529	2	-19.3	-4.0	6.4	2.4	-5.9	-6.3	-0.4	-16.9
ZINC00076824	2	-18.9	-4.2	6.2	2.0	-6.9	-7.3	-0.4	-16.9
ZINC04940173	2	-19.0	-3.8	6.0	2.2	-4.7	-6.1	-1.4	-16.9
ZINC14505663	2	-16.1	-8.2	7.5	-0.7	-7.3	-10.7	-3.4	-16.8
ZINC02732051	2	-19.2	-5.1	7.5	2.5	-5.7	-6.1	-0.4	-16.8
ZINC45236747	2	-18.2	-4.8	6.3	1.5	-4.1	-5.6	-1.5	-16.7
ZINC36756146	2	-17.7	-4.6	5.5	0.9	-5.3	-6.5	-1.3	-16.7
ZINC35207533	2	-17.3	-3.4	4.4	1.0	-5.1	-6.4	-1.3	-16.3
ZINC12783212	2	-21.1	-8.3	13.2	4.9	-6.3	-6.4	-0.1	-16.1
ZINC59919622	2	-17.7	-5.3	6.9	1.7	-6.2	-7.1	-0.8	-16.0
ZINC31439120	2	-20.7	-8.4	13.0	4.6	-6.5	-6.6	-0.1	-16.0
ZINC94618071	2	-17.4	-9.6	13.1	3.5	-7.5	-8.0	-0.5	-13.9
ZINC01089842	2	-13.8	-4.9	5.4	0.5	-6.7	-7.3	-0.6	-13.3
ZINC01679277	2	-15.6	-4.4	6.8	2.5	-6.2	-6.4	-0.2	-13.2
ZINC71920814	3	-25.6	-10.8	15.5	4.7	-9.1	-9.2	-0.2	-20.9
ZINC40773700	3	-21.0	-12.1	16.2	4.1	-8.3	-9.7	-1.3	-17.0
ZINC31439120	3	-20.7	-8.4	12.9	4.5	-6.5	-6.6	-0.1	-16.1
ZINC37864247	3	-25.1	-5.5	17.5	12.1	-41.3	-9.8	31.5	-13.0
ZINC33269125	4	-29.8	-12.1	19.9	7.7	-35.9	-10.1	25.8	-22.0
ZINC12769485	4	-27.9	-12.9	19.3	6.4	-32.8	-11.0	21.8	-21.5
ZINC95449202	4	-22.6	-4.6	13.7	9.1	-9.0	-3.4	5.6	-13.6
ZINC04181101	4	-23.2	0.3	10.6	11.0	-17.5	-5.1	12.4	-12.3

Table 1

Figure and Table legends:

Figure 1: Selection of small molecules for which inhibitory and dissociation constants have been reported for SMARCA4. **a,b)** Four derivatives of a single compound developed for BET bromodomains. **c)** 1,3-dimethyl benzimidazole compound developed for BRPF1 pK_D value obtained for SMARCA4 through via bromoscan. **d)** Only PFI3 is reported as a dedicated chemical probe for SMARCA4 while other reported inhibitors were identified as part of investigations into selectivity for the original target. All 2D depictions of small molecules were generated via ChemDraw Pro 12 (CambridgeSoft, Cambridge, MA, United States).

Figure 2: Overview of SMARCA4 sequence and structural information. **a)** Annotated sequence of the SMARCA4 bromodomain. Fragments were docked to the residues underlined in black. **b)** Three dimensional structure of SMARCA and 1-methylpyrrolidin-2-one forming a single hydrogen bond with Asn1540 (PDB-ID: 3UVD). The motifs of the conserved bromodomain fold are colored according to subplot a). Structural waters and 1-methylpyrrolidin-2-one are shown in sticks. **c)** Kac binding interface of SMARCA4. Residues involved in fragment and molecule docking are shown in sticks and annotated. Residues highlighted in had a conformation providing hydrogen bond acceptors while residues highlighted in purple could contribute both H-bond acceptor and donor. All other residues only engaged in hydrophobic interactions with a potential ligand. The majority of residues in the binding site are of hydrophobic nature. Figures b) and c) rendered in Pymol Version 1.7.4 Schrödinger, LLC.

Figure 3: Overview of the virtual screening with annotated number of molecules and fragments respectively. Each step in the virtual screening for the molecule stage generally features the desired reduction of chemical space by a factor of 10. The fragment stage is shown in dashed lines while the full molecule stage is shown in solid lines. The only compound showing activity 20 μM is depicted in 2D.

Figure 4: IC_{50} curves for the control PFI3 and ZINC04181101 measured via ALPHA screen.

Figure 5: Docked pose of ZINC04181101. The hydrogen annotated in yellow dashed lines show two hydrogen bonds formed with Asn1540 while a third one is formed with the backbone of Gln1487. The trans-glycoluril motif contributes a total of 2 hydrogen bonds both as acceptor and donor. In this pose the phenol group does not feature any distinct interaction. No hydrogen bonds with water were detected. Waters, side chain atoms, and backbone atoms for Gln1487 depicted as sticks. Figure generated in Pymol[59].

Table 1: Table of the 32 tested compounds and their respective *in silico* filter criteria after tethered docking. Before CHARMM energy evaluation the number of hydrogen bonds was the prime filtering criterion (column 2). Following CHARMM energy evaluation all molecules with a negative E_{total} ($E_{vdw} + \Delta G_{elec}$) were discarded (column 10). ΔG_{elec} is calculated as the sum of $\Delta E_{Coulomb}$ and ΔE_{solv}. ΔElect was devised as a measure to assess favorable electrostatic interactions of the ligand inside the receptor compared to its solvated state and is calculated as the difference between $E_{inter,solv}$ and $\Delta E_{solv,lig}$. Molecules featuring 3 or 4 hydrogen bonds have a comparatively high ΔElect as their scaffolds provided more hydrogen bond donors/acceptors.

References:

1. Dawson MA, Kouzarides T, Huntly BJP. Targeting Epigenetic Readers in Cancer. *N Engl J Med*. 2012;367(7):647-657. doi:10.1056/NEJMra1112635.

2. Virani S, Colacino JA, Kim JH, Rozek LS. Cancer Epigenetics: A Brief Review. *ILAR J*. 2012;53(3-4):359-369. doi:10.1093/ilar.53.3-4.359.

3. Roy DM, Walsh LA, Chan TA. Driver mutations of cancer epigenomes. *Protein Cell*. 2014;5(4):265-296. doi:10.1007/s13238-014-0031-6.

4. Barbieri I, Cannizzaro E, Dawson MA. Bromodomains as therapeutic targets in cancer. *Brief Funct Genomics*. 2013;12(3):219-230. doi:10.1093/bfgp/elt007.

5. Muller S, Filippakopoulos P, Knapp S. Bromodomains as therapeutic targets. *Expert Rev Mol Med*. 2011;13. doi:10.1017/S1462399411001992.

6. Papavassiliou KA, Papavassiliou AG. Bromodomains: pockets with therapeutic potential. *Trends Mol Med*. 2014;20(9):477-478. doi:10.1016/j.molmed.2014.06.004.

7. Delmore JE, Issa GC, Lemieux ME, et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*. 2011;146(6):904-917. doi:10.1016/j.cell.2011.08.017.

8. Filippakopoulos P, Knapp S. The bromodomain interaction module. *FEBS Lett*. 2012;586(17):2692-2704. doi:10.1016/j.febslet.2012.04.045.

9. Trotter KW, Archer TK. The BRG1 transcriptional coregulator. *Nucl Recept Signal*. 2008;6. doi:10.1621/nrs.06004.

10. Wong AKC, Shanahan F, Chen Y, et al. BRG1, a Component of the SWI-SNF Complex, Is Mutated in Multiple Human Tumor Cell Lines. *Cancer Res*. 2000;60(21):6171-6177.

11. Medina PP, Romero OA, Kohno T, et al. Frequent BRG1/SMARCA4–inactivating mutations in human lung cancer cell lines. *Hum Mutat*. 2008;29(5):617-622. doi:10.1002/humu.20730.

12. Medina PP, Sanchez-Cespedes M. Involvement of the chromatin-remodeling factor BRG1/SMARCA4 in human cancer. *Epigenetics*. 2008;3(2):64-68. doi:10.4161/epi.3.2.6153.

13. Tsurusaki Y, Okamoto N, Ohashi H, et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet*. 2012;44(4):376-378. doi:10.1038/ng.2219.

14. Buscarlet M, Krasteva V, Ho L, et al. Essential role of BRG, the ATPase subunit of BAF chromatin remodeling complexes, in leukemia maintenance. *Blood*. 2014;123(11):1720-1728. doi:10.1182/blood-2013-02-483495.

15. Witkowski L, Carrot-Zhang J, Albrecht S, et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat Genet*. 2014;46(5):438-443. doi:10.1038/ng.2931.

16. Moes-Sosnowska J, Szafron L, Nowakowska D, et al. Germline SMARCA4 mutations in patients with ovarian small cell carcinoma of hypercalcemic type. *Orphanet J Rare Dis*. 2015;10(1):32. doi:10.1186/s13023-015-0247-4.

17. Hohmann AF, Vakoc CR. A rationale to target the SWI/SNF complex for cancer therapy. *Trends Genet*. 2014;30(8):356-363. doi:10.1016/j.tig.2014.05.001.

18. Wang X, Sansam CG, Thom CS, et al. Oncogenesis Caused by Loss of the SNF5 Tumor Suppressor Is Dependent on Activity of BRG1, the ATPase of the SWI/SNF Chromatin Remodeling Complex. *Cancer Res*. 2009;69(20):8094-8101. doi:10.1158/0008-5472.CAN-09-0733.

19. Holik AZ, Young M, Krzystyniak J, et al. Brg1 Loss Attenuates Aberrant Wnt-Signalling and Prevents Wnt-Dependent Tumourigenesis in the Murine Small Intestine. *PLoS Genet*. 2014;10(7):e1004453. doi:10.1371/journal.pgen.1004453.

20. Shen W, Xu C, Huang W, et al. Solution Structure of Human Brg1 Bromodomain and Its Specific Binding to Acetylated Histone Tails†,‡. *Biochemistry (Mosc)*. 2007;46(8):2100-2110. doi:10.1021/bi0611208.

21. Zhang G, Plotnikov AN, Rusinova E, et al. Structure-Guided Design of Potent Diazobenzene Inhibitors for the BET Bromodomains. *J Med Chem*. 2013;56(22):9251-9264. doi:10.1021/jm401334s.

22. Demont EH, Bamborough P, Chung C, et al. 1,3-Dimethyl Benzimidazolones Are Potent, Selective Inhibitors of the BRPF1 Bromodomain. *ACS Med Chem Lett*. 2014;5(11):1190-1195. doi:10.1021/ml5002932.

23. Picaud S, Wells C, Felletar I, et al. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc Natl Acad Sci*. 2013;110(49):19754-19759. doi:10.1073/pnas.1310658110.

24. Ciceri P, Müller S, O'Mahony A, et al. Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. *Nat Chem Biol*. 2014;10(4):305-312. doi:10.1038/nchembio.1471.

25. Zhao L, Wang Y, Cao D, et al. Fragment-Based Drug Discovery of 2-Thiazolidinones as BRD4 Inhibitors: 2. Structure-Based Optimization. *J Med Chem*. 2015;58(3):1281-1297. doi:10.1021/jm501504k.

26. PFI-3: Selective chemical probe for SMARCA bromodomains. http://www.thesgc.org/chemical-probes/PFI-3.

27. Kolb P, Kipouros CB, Huang D, Caflisch A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins*. 2008;73(1):11-18. doi:10.1002/prot.22028.

28. Frances C. Bernstein, Thomas F. Koetzle, Graheme J.B. Williams, Edgar F. Meyer, Michael D. Brice, John R. Rodgers, Olga Kennard, Takehiko Shimanouchi MT. The Protein D a t a Bank : A Computer-based Archival File for Macromolecular Structures The Protein Data Bank is a computer-based archival file for macromolecular (a) Scope. *J Mol Biol*. 1977;(112):535-542.

29. Filippakopoulos P, Picaud S, Mangos M, et al. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell*. 2012;149(1):214-231. doi:10.1016/j.cell.2012.02.013.

30. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-38, 27-28.

31. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem*. 2013;34(25):2135-2145. doi:10.1002/jcc.23354.

32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926. doi:10.1063/1.445869.

33. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983;4(2):187-217. doi:10.1002/jcc.540040211.

34. Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: The biomolecular simulation program. *J Comput Chem*. 2009;30(10):1545-1614. doi:10.1002/jcc.21287.

35. MacKerell AD, Brooks B, Brooks CL, et al. CHARMM: The Energy Function and Its Parameterization. In: *Encyclopedia of Computational Chemistry*. Vol John Wiley & Sons, Ltd; 2002. http://onlinelibrary.wiley.com/doi/10.1002/0470845015.cfa007/abstract. Accessed March 28, 2015.

36. Irwin JJ, Shoichet BK. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. *J Chem Inf Model*. 2004;45(1):177-182. doi:10.1021/ci049714+.

37. Kolb P, Caflisch A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J Med Chem*. 2006;49(25):7384-7392. doi:10.1021/jm060838i.

38. Yesselman JD, Price DJ, Knight JL, Brooks CL. MATCH: An atom-typing toolset for molecular mechanics force fields. *J Comput Chem*. 2012;33(2):189-202. doi:10.1002/jcc.21963.

39. Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010;31(4):671-690. doi:10.1002/jcc.21367.

40. Meslamani J, Smith SG, Sanchez R, Zhou M-M. ChEpiMod: a knowledgebase for chemical modulators of epigenome reader domains. *Bioinformatics*. 2014;30(10):1481-1483. doi:10.1093/bioinformatics/btu052.

41. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins Struct Funct Bioinforma*. 1999;37(1):88-105. doi:10.1002/(SICI)1097-0134(19991001)37:1<88::AID-PROT9>3.0.CO;2-O.

42. Landrum G. RDKit: Open-source cheminformatics. http://www.rdkit.org/.

43. Hann M, Hudson B, Lewell X, Lifely R, Miller L, Ramsden N. Strategic Pooling of Compounds for High-Throughput Screening. 1999:897-902.

44. Brenk R, Schipani A, James D, et al. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*. 2008;3(3):435-444. doi:10.1002/cmdc.200700139.

45. Morley SD, Afshar M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®. *J Comput Aided Mol Des*. 2004;18(3):189-208. doi:10.1023/B:JCAM.0000035199.48747.1e.

46. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, et al. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput Biol*. 2014;10(4):e1003571. doi:10.1371/journal.pcbi.1003571.

47. Clark M, Cramer RD, Van Opdenbosch N. Validation of the general purpose tripos 5.2 force field. *J Comput Chem*. 1989;10(8):982-1012. doi:10.1002/jcc.540100804.

48. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminformatics*. 2011;3(1):33. doi:10.1186/1758-2946-3-33.

49. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J Chem Inf Model*. 2012;52(12):3144-3154. doi:10.1021/ci300363c.

50. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J Chem Inf Model*. 2012;52(12):3155-3168. doi:10.1021/ci3003649.

51. Im W, Beglov D, Roux B. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput Phys Commun*. 1998;111(1–3):59-75. doi:10.1016/S0010-4655(98)00016-2.

52. Berthold MR, Cebron N, Dill F, et al. KNIME: The Konstanz Information Miner. In: Preisach C, Burkhardt PDH, Schmidt-Thieme PDL, Decker PDR, eds. *Data Analysis, Machine Learning and Applications*. Vol Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg; 2008:319-326. http://link.springer.com/chapter/10.1007/978-3-540-78246-9_38. Accessed May 17, 2015.

53. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J Chem Inf Comput Sci*. 1989;29(3):163-172. doi:10.1021/ci00063a006.

54. Wildman SA, Crippen GM. Prediction of Physicochemical Parameters by Atomic Contributions. *J Chem Inf Comput Sci*. 1999;39(5):868-873. doi:10.1021/ci990307l.

55. Zhang G, Sanchez R, Zhou M-M. Scaling the Druggability Landscape of Human Bromodomains, a New Class of Drug Targets. *J Med Chem*. 2012;55(17):7342-7345. doi:10.1021/jm3011977.

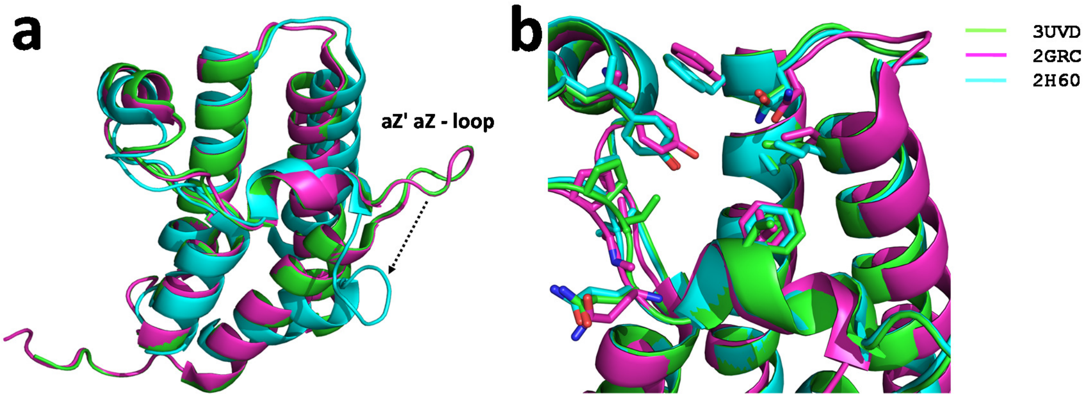
56. Vidler LR, Brown N, Knapp S, Hoelder S. Druggability Analysis and Structural Classification of Bromodomain Acetyl-lysine Binding Sites. *J Med Chem*. 2012;55(17):7346-7359. doi:10.1021/jm300346w.

57. Filippakopoulos P, Knapp S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov*. 2014;13(5):337-356. doi:10.1038/nrd4286.

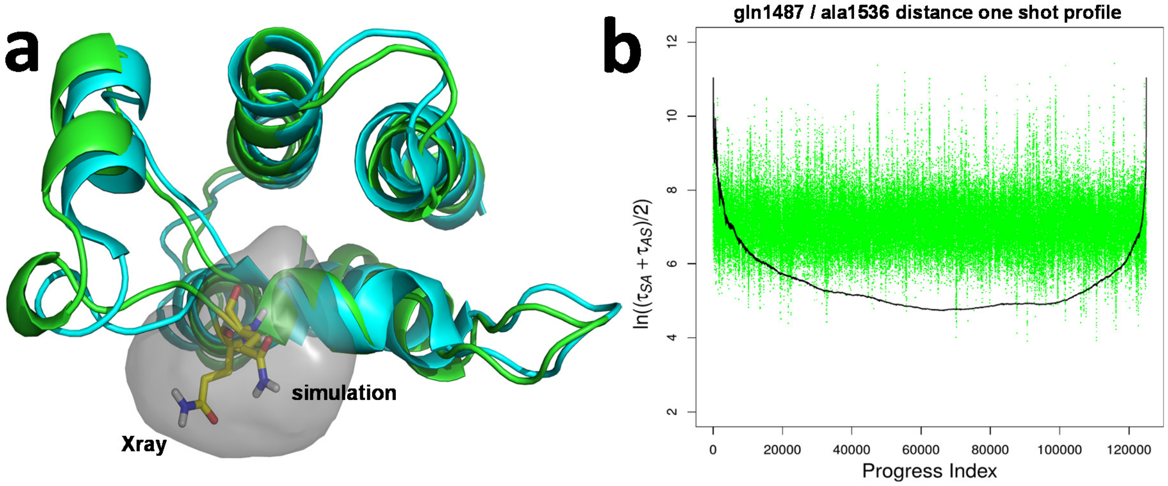
58. The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

59. Steiner S, Magno A, Huang D, Caflisch A. Does bromodomain flexibility influence histone recognition? *FEBS Lett*. 2013;587(14):2158-2163. doi:10.1016/j.febslet.2013.05.032.

Supplementary Material:



Supplementary 1: Alignment of the available crystal structures. **a)** 3UVD and 3GRC display very little difference as the RMSD is 0.27 Å while the RMSD between 3UVD and 2HR6 (NMR) is 1.85 Å. The biggest difference between the two structures is a different orientation of the aZ’-aZ loop (black dashed arrow). b) Close-up of the aligned binding sites of the respective SMARCA4 structures reveals that there are little differences in the residue orientations between the structures.



Supplementary 2: Flexibility of GLN1487. **a)** Surface view for all primary amine nitrogen atoms from the simulation indicating high flexibility and overall movement of the sidechain in the simulation performed by Steiner et al.[52] b) Progress index (black line) and annotated distances between Gln1487 sidechain amide nitrogen. The progress index shows a single basin indicating no stable or metastable conformations of Gln1487. The progress index was generated in CAMPARI by involving all side chain heavy atoms of GLN1487. Prior to clustering via the tree-based cluster algorithm implemented in CAMPARI the backbone of all SMARCA4 residues were aligned.

```
[Br,Cl,I][CX;CH,CH2], reactive alkyl halides
[S,C](=[O,S])[F,Br,Cl,I] acid halides
O=CN=[N+]=[N-] carbazides
COS(=O)=[C,c] sulphate esters
C(=O)OC(=O) acid anhydrides
OO peroxides
C(=O)Oc1c(F)c(F)c(F)c(F)c1(F) pentafluorophenyl esters
C(=O)Oc1ccc(C(=O)=O)cc1 paranitrophenyl esters
C(=O)Onnn esters of HOBT
N=C=[S,O] Isocyanates Isothiocyanates
OS(=O)(=O)C(F)(F)F Triflates
P(=S)(S)S lawesson's reagent and derivatives
cN=[N+]=[N-] aromatic azides
C(=O)C[N+,n+] beta carbonyl quaternary nitrogen
C=P posphoranes
[Cl]C([C&R0])=N chloramidines
[N&D2](=O) nitroso
[P,S][Cl,Br,F,I] P/S Halides
N=C=N Carbodiimide
C(=O)N(C(=O))OC(=O) triacyloximes
N#CC[OH] cyanohydrins
P(OCC)(OCC)(=O)C#N cyanophophonates
C(=O)O[C,H1].C(=O)O[C,H1].C(=O)O[C,H1] > 2 ester groups
C(=O)[Cl,Br,I,F] acid halide
[CH1](=O) aldehyde
N=[N+]=[N-] Azido group
N#N Azo group
[C,c](=O)[CX4,CR0X3,O][C,c](=O) beta-keto/anhydride
[C+,c+,C-,c-] Carbo cation/ anion
[O+,o+,S+,s+] charged oxygen or sulfur atoms
C1(=[O,N])C=CC(=[O,N])C=C1 chinone
C1(=[O,N])C(=[O,N])C=CC=C1 chinone
C=[C!r]C#N conjugated nitrile group
[N!R]=[N!R] diazo group
[C,c](=O)[C,c](=O) diketo group
SS disulphide
C(=O)Onnn ester of HOBT
[Hg,Fe,As,Sb,Zn,Se,se,Te,B,Si] heavy metal
[NX3R0,NX4R0,OR0,SX2R0][CX4][NX3R0,NX4R0,OR0,SX2R0] het-C-het not in ring
N[NH2] hydrazine
C=[N!R] imine
N=[CR0][N,n,O,S] imine
N=C=O isocyanate
C=C=O ketene
S1C=CSC1=S methylidene-1,3-dithiole
C=!@CC=[O,S] Michael acceptor
[$([CH]),$(CC)]#CC(=O)[C,c] Michael acceptor
[$([CH]),$(CC)]#CS(=O)(=O)[C,c] Michael acceptor
C=C(C=O)C=O Michael acceptor
[$([CH]),$(CC)]#CC(=O)O[C,c] Michael acceptor
[NX2,nX3][OX1] N oxide
[NX3,NX4][F,Cl,Br,I] N-halo
n[OH] N-hydroxyl pyridine
[#7]-N=O N-nitroso
[C,c]=N[OH] oxime
[C,c]=NOC=O oxime
OO peroxide
c1ccccc1OC(=O)[#6] phenol ester
c1ccccc1OC(=O)O phenyl carbonate
[CR0]=[CR0][CR0]=[CR0] polyene
[s,S,c,C,n,N,o,O]~[nX3+,NX3+](~[s,S,c,C,n,N])~[s,S,c,C,n,N] quaternary nitrogen
[s,S,c,C,n,N,o,O]~[n+,N+](~[s,S,c,C,n,N,o,O])(~[s,S,c,C,n,N,o,O])~[s,S,c,C,n,N,o,O]  quaternary nitrogen
[*]=[N+]=[*] quaternary nitrogen
[Si][F,Cl,Br,I] silicon halogen
[SX3](=O)[O-,OH] sulfinic acid
[SX2]O sulfur oxygen single bond
SC=O thioester
*1[O,S,N]*1 Three-membered heterocycle
OS(=O)(=O)C(F)(F)F triflate
[SiR0,CR0](c1ccccc1)(c2ccccc2)(c3ccccc3) triphenyl methyl-silyl
```

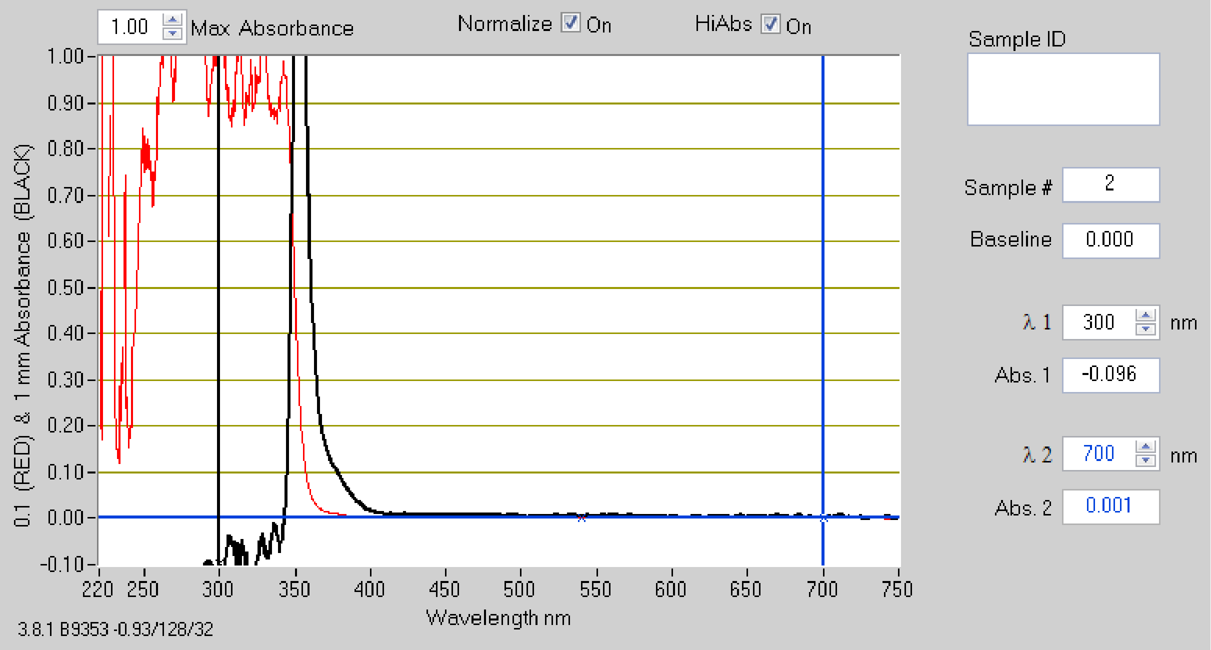
Supplementary 3: SMARTS patterns for the reduction of chemical space.

ZINC_ID	measurement 1	measurement 2	average
ZINC95356471	104.70	104.01	104.4
ZINC04206597	104.01	101.91	103.0
ZINC41121701	104.25	100.45	102.4
ZINC22642290	91.97	92.32	92.1
ZINC95449202	110.60	105.66	108.1
ZINC01089842	102.48	103.56	103.0
ZINC02732051	98.70	102.05	100.4
ZINC05417915	106.42	104.11	105.3
ZINC04181101	14.44	14.38	14.4
ZINC33269125	106.43	103.64	105.0
ZINC04940173	96.69	99.30	98.0
ZINC04940185	94.11	94.16	94.1
ZINC00127673	103.70	100.90	102.3
ZINC00076820	97.68	100.17	98.9
ZINC00076824	99.37	98.11	98.7
ZINC31439120	101.83	99.76	100.8
ZINC04206529	99.10	104.50	101.8
ZINC36756146	97.41	105.80	101.6
ZINC12783212	104.46	102.54	103.5
ZINC12769485	89.68	90.51	90.1
ZINC12901687	87.93	84.81	86.4
ZINC33207842	95.71	95.06	95.4
ZINC12858178	99.64	102.74	101.2
bromosporine	46.57	46.12	46.3
ZINC42735712	101.83	101.77	101.8
ZINC12863354	79.05	78.48	78.8
ZINC84678055	96.01	95.42	95.7
ZINC32855766	93.78	91.62	92.7
ZINC14121148	88.86	90.34	89.6
ZINC12804112	97.48	97.83	97.7
ZINC66699837	104.93	102.60	103.8
ZINC05417915	110.52	114.46	112.5
ZINC71920814	98.60	98.11	98.4

Supplementary 4: Single dose ALPHA screen measurements. The competition-based assay results in a reduced when K14/18/23/27Ac-GG is displaced by the respective inhibitor. The values represent percentages in change of fluorescence at 520-620 nm in comparison to a solution with no inhibitor present.

Raw Data	Conc (M)	S9	PFI3	Conc (M)
	1.00E-04	23995	16277	**4.00E-05**
	3.33E-05	78394	17787	**1.33E-05**
	1.11E-05	154392	24740	**4.44E-06**
	3.70E-06	177654	39825	**1.48E-06**
	1.23E-06	212296	82876	**4.94E-07**
	4.12E-07	214575	138648	**1.65E-07**
	1.37E-07	210676	188650	**5.49E-08**
	4.57E-08	211693	220931	**1.83E-08**
	1.52E-08	222241	226555	**6.10E-09**
	5.08E-09	214781	221497	**2.03E-09**
	DMSO	224381	230808	**DMSO**

Supplementary 5: Raw data of ALPHA Screen measurements for the IC_{50} curves of PFI3 and ZINC04181101 (S9:internal naming).



Max Absorbance 1.00

Normalize ☑ On

HiAbs ☑ On

Sample ID

Sample # 2

Baseline 0.000

λ 1 300 nm

Abs. 1 -0.096

λ 2 700 nm

Abs. 2 0.001

Supplementary 6: Absorption spectrum for ZINC04181101 diluted to 50 µM in DMSO. The measurement was performed on a Nanodrop™ in a wavelength region between 300 and 700 nm against a blank solution of DMSO.

Chapter 5

Molecular dynamics analysis of iriomoteolide 3a and reidispongiolide A binding to monomeric actin

5.1 Introduction

5.1.1 The cytoskeleton

The cytoskeleton is constituted of three prime cytoskeletal structures: microtubules (MT), intermediary filaments (IF) and actin/micro filaments (MF) (Figure 1)[1–3]. In the cell these structures are present as polymers of varying lengths, constituted of highly ordered and repetitive units of various structural proteins, most prominently actin, tubulin, and the sequence homology class (SHC) members of IF.

Microtubules are cylindrical, hollow polymers with an outer diameter of 24 nm and an inner diameter of 12 nm composed of different tubulin species (α, β, γ, δ, ε, ζ) featuring several subtypes each (Figure 2). Structure and composition are generally determined by cell type and functional role of the microtubule. The most common tubulin species are α- and β- tubulin which respectively have a molecular weight of ~ 50 kDa. Via their GTPase activity, α- and β- tubulin assemble initially by forming αβ-hetero dimers which in-turn assemble, head to tail, into protofilaments. An average of 13 protofilaments assemble into short-lived sheets which then laterally fold into the final cylindrical and polar structure with a fast-growing plus-end (+) and a slow growing minus-end (-)[4]. In eukaryotes the minus-end is generally attached at the centrosomes, where γ-tubulin plays a key role initializing protofilament formation, while the (+) end grows towards the designated cell compartment. Microtubules have been shown to stochastically undergo phases of shrinking and growing *in vitro* and *in vivo*[5–7]. This dynamic instability is central to the function of the microtubule as this, for example, is the mechanism by which chromosomes are separated upon mitosis[8]. The dynamics of rescue (growth after a pause event) and catastrophe (shrinkage after a pause event) are influenced by a multitude of proteins as well as organic- and inorganic molecules[9]. The polymerization of αβ-hetero dimers into protofilaments requires the energy set free upon hydrolysis of guanosin-5'-triphosphate (GTP) bound to β-tubulin into remaining guannosin-5'-diphosphate (GDP) and dissociating inorganic phosphate[10]. The specific arrangement of an inter-dimer interface is crucial for the GTPase activity which leads to hydrolysis and subsequent polymerization. Polymerization of the tubule is halted by GTP caps which prevent further polymerization and depolymerization[11,12]. Residual GTPs bound to β- tubulins are also distributed within the fiber which is hypothesized to be integral in the dynamic equilibrium between polymerization and depolymerization of the microtubule. The dynamic changes in length of

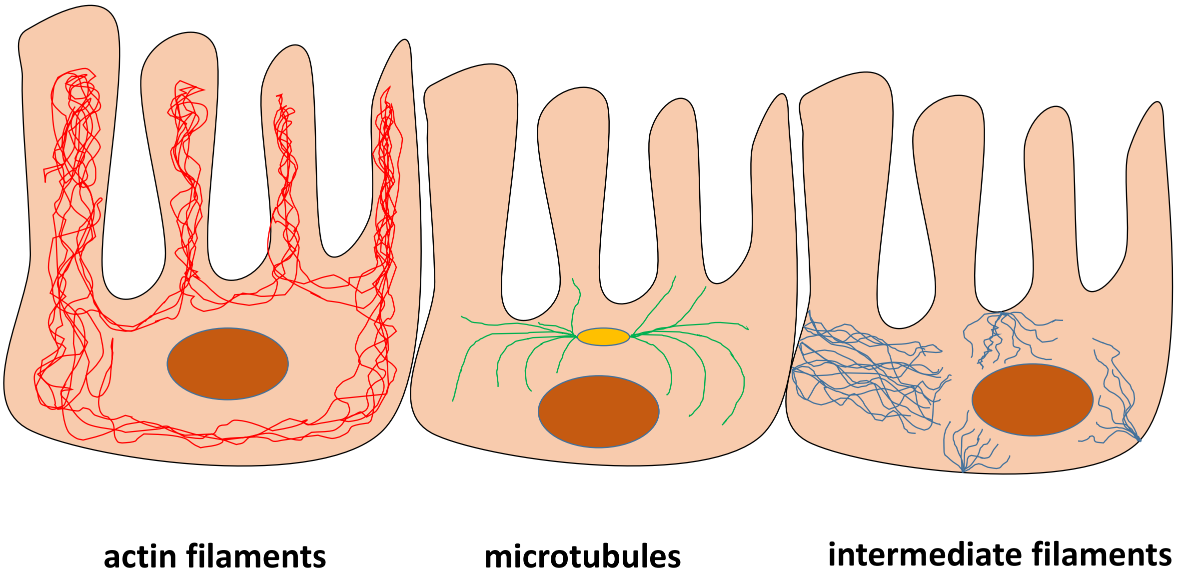


Figure 1: Schematic of epithelial cells depicting the three main constituents of the cytoskeleton. Actin filaments/micro filaments are depicted in red while microtubules and intermediate filaments are depicted in green and blue respectively. The nucleus is highlighted as a brown ellipsoid while the spindle apparatus is shown in yellow.

microtubules is central in the relocation of cell compartments and multitudes of proteins[13]. Microtubules feature one of the largest amount of protein-protein interactions. Not only are proteins influencing polymerization and depolymerization rates of microtubules, they also associate and re-locate alongside the stable tubule regions.

Compared to microtubules, intermediate filaments (IF) are made up of a vast number of different proteins and are usually located in the cytoplasm as well as the nucleus[14,15]. Their prime function in adult organisms is the shaping of cells and their interfaces the adjacent environment. IF are categorized into 5 sequence homology classes (SHC)[16]. Acidic and basic keratins are type 1 and 2. Type 3 encompasses vimentin, desmin, and GFAP while neurofilament proteins are type 4 and nuclear lamins make up type 5 IF[17]. Sequence and structural analysis has revealed that there are several proteins (e.g nestin and filesin) which are highly similar to the IF superfamily but can't be directly categorized into the above five classes. Despite these five different types and several subtypes, all IFs and related proteins share a similar 45 nm central α-helical 'rod'-domain encompassed by N-terminal (head) and C-terminal (tail) domains (Figure 3)[16]. Notably, the assembly into the final cytoskeletal structure happens via lateral association of the subunits within seconds. Additionally, the assembly doesn't require co-factors compared to microtubules or microfilaments and the final IF is highly robust against a multitude of environmental factors like pH, temperature, and mechanical stress.

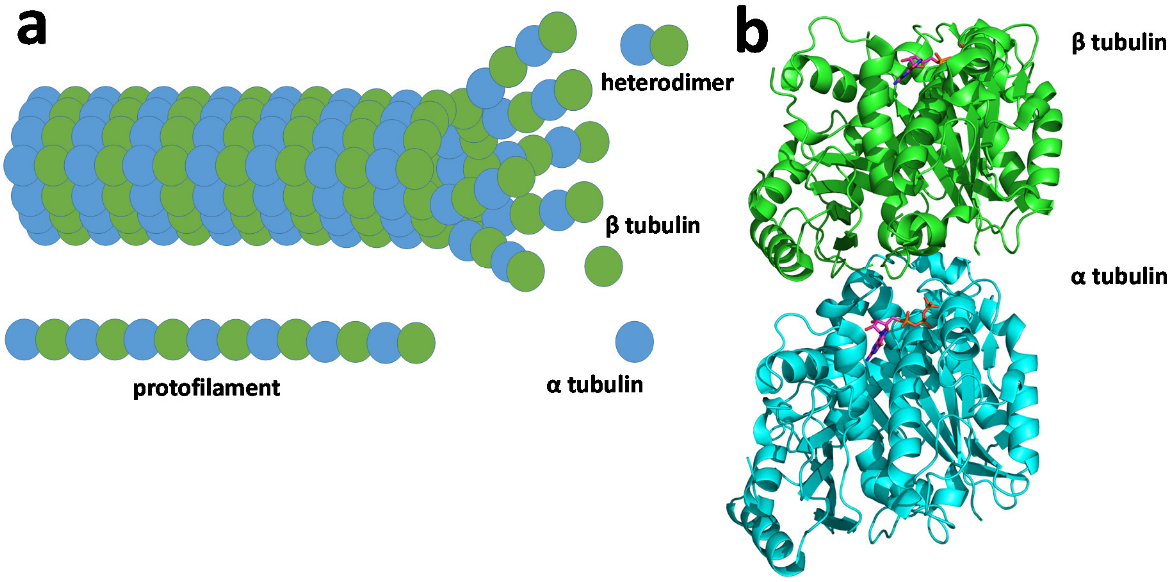


Figure 2: Schematic of microtubule assembly. **a)** Intermediaries of the microtubules including the two subunits α,β tubulin and an example protofilament of which several are assembled into the final microtubule. **b)** Pymol-generated depiction of a X-ray crystal structure of the α,β tubulin dimer. GTP is shown in magenta sticks. (PDB-ID: 4FFB, protein STU2 and Mg^{2+} are omitted)[18]. All Pymol-related images were generated via version 1.7.0.0[19].

At the beginning of the assembly, two polypeptide chains align in parallel and in register to form the initial coiled coil complex which further matures into the IF structure at low pH with the addition of other proteins (Figure 3)[20]. The alignment happens without the assistance of auxiliary proteins and is therefore considered intrinsic to the structure. IF notably are not made up of globular subunits but long polypeptides aligned in an anti-parallel manner resulting in a non-polar super structure. The most distinguishing feature of IF, is the ability of high extensibility and stress resistance not observed in MT or MF. This property is related to the coiled coil structure of the filaments. Diseases related to mutations in IF protein genes like keratin present themselves in severe structural weaknesses of cells, tissues, and organs at very distinct stages of development. Muscular dystrophies, cardiovascular as well as Hutchinson-Gilford progeria syndrome and atypical Werner syndrome occur in post-infant stages while mutations in GFAP genes lead to severe disruption of normal development of the brain and skull (Alexander disease)[21]. Epidermolytic diseases become less severe with age, indicating a change of molecules interacting with IF leading to hypotheses that such diseases result in the hindered interplay of proteins associated to IF.

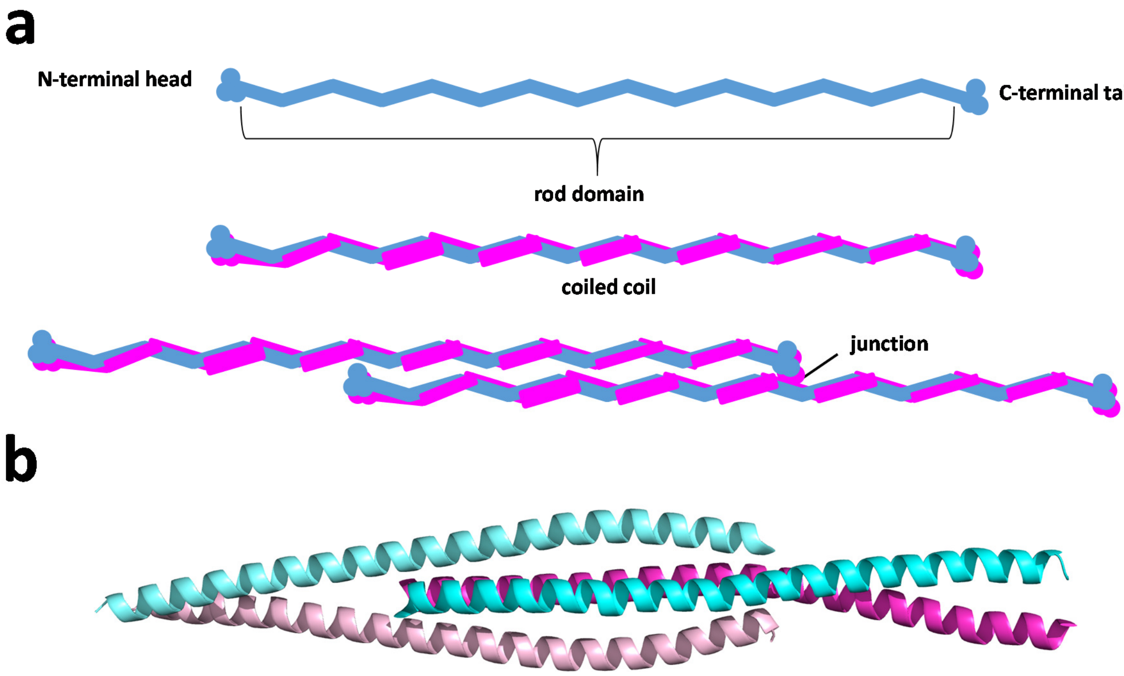


Figure 3: Simplified overview intermediate filament assembly. **a)** Example schematic of stages of IF assembly. From top to bottom a simple monomer, a coiled coil dimer, and an associated tetramer stemming from one dimer associated via the C-terminal tail to the coil of an adjacent dimer. **b)** X-ray crystal structure of a fragment of a vimentin coil 2 homo 4-mer (PDB-ID: 3KLT)[22]. Each individual chain in shown in a different color. All non-peptidic content was removed before rendering the image with Pymol[19].

5.1.2 Actin

The dynamic structure of actin filaments is vital for every biological mechanism that requires cellular scaffolding. As one of the most abundant proteins in eukaryotic cells (ranging from 5% - 10% in any cell to ~20% in muscle cells), actin filaments provide and maintain the cellular three-dimensional structure, provide cell motility, are vital in cytokinesis, and muscle contraction[23]. Actin is most abundantly located in the near cell membrane region called the cell cortex. Compared to microtubules and intermediate filaments, the length of actin filaments are the shortest with a persistence length of 3-6 nm. It's dynamics of polymerization/depolymerization as well as the ability of crosslinking enable actin towards a broad diversity of mechanical behaviors. Actin is highly conserved among species with an average of 80% identity between humans and yeast and has many related proteins in prokaryotes[24]. Myriads of proteins are known to bind to actin of which myosin is probably the most prominent one, determining muscle contraction and relaxation[25]. Proteins binding to actin are not only subject to its relocating effects but also regulate polymerization and depolymerization as well as crosslinking[25].

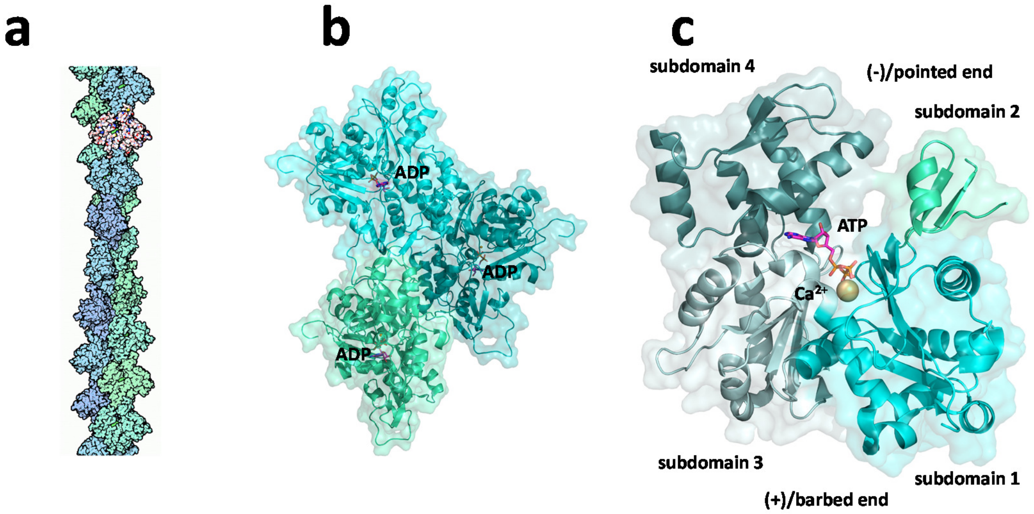


Figure 4: Systematic close up of an actin filament towards the monomeric actin subunit. **a)** Schematic of an actin filament (July 2001 Molecule of the Month by David Goodsell, RSCB Protein Data Bank, www.rcsb.org). **b)** Hypothetical model of three actin monomers with ADP as the bound nucleotide and dummy ion as a substitute for Mg^{2+}/Ca^{2+} [26]. Each monomer is represented as cartoon in a distinctly colored cyan, the molecular surface is kept transparent. Image rendered in Pymol. **c)** Pymol-visualized monomeric G-actin with bound ATP and Ca^{2+} (PDB-ID: 2HF4). The subunits 1-4 are colored in distinct shades of cyan while the nucleotide is shown in magenta sticks and the Ca^{2+} ion is represented by an orange sphere[19].

Additionally to proteins, many small organic/inorganic molecules have been identified to bind to actin with ATP being the co-factor which enables polymerization of the globular G-Actin into F-Actin while other known non-protein molecules have an adverse effect on innate actin dynamics[27].

5.1.3 Monomeric globular actin and fibrous actin

Vertebrates encode genes for three main globular actin (G-actin) isoforms α-, β-, and γ-actin which are categorized by their unique isoelectric points. These isoforms are further differentiated into $\alpha_{skeletal}$-actin, $\alpha_{cardiac}$-actin, α_{smooth}-actin and $\gamma_{skeletal}$-actin which are primarily expressed in skeletal, cardiac, and smooth muscle tissue. γ_{cyto}-actin and β_{cyto}-actin and expressed equally in all tissues[28]. High conservation of all six isoforms is present between mammals in birds where 93% sequence identity represents the low margin. Sequence differences are the highest towards the N-terminal region between different isoforms and their homologues[29]. The G-actin monomer has a molecular weight of 43 kDa comprised of 375 residues. Flattened in one dimension, actin is made up of two major domains, a large inner domain and a small outer domain. These domains are further annotated into four subdomains (domains 1, 2, 3, 4) with the nucleotide binding cleft (NBC) located at the center between subdomains

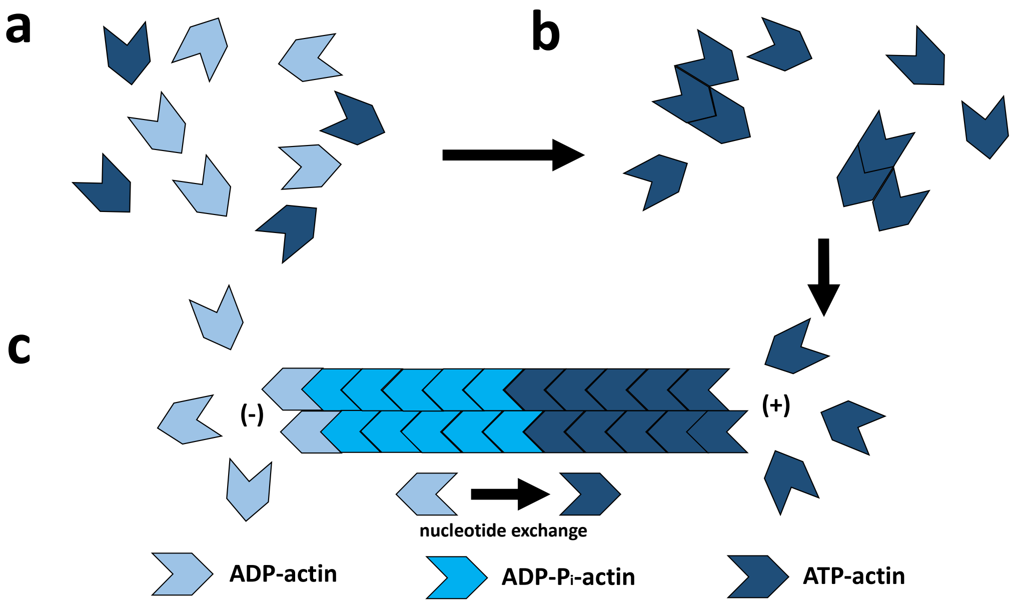


Figure 5: Overview of actin polymerization/depolymerization dynamics. **a)** Mixture of ADP-actin and ADP-P_i-actin. Without a specific environment monomeric actin will not associate beyond the stage of loose oligomers. **b)** The formation of stable nuclei is induced at high salt concentrations. Loose oligomers from which do not serve as stable seeds are associate regardless. **c)** Stable nuclei are polymerized into an actin filament. Polymerization and depolymerization takes place at both barbed end and pointed end. Nucleotide exchange from ADP-actin to ATP-actin is what enables tread milling of the filament. All steps are influenced by a multitude of environmental factors such as pH, salt concentration, proteins, and organic/inorganic molecules.

1,2 and 3,4 (Figure 4c)[30]. Together with the respective nucleotides adenosine-tri-phosphate (ATP)/adenosine-di-phosphate (ADP), a divalent cation Ca^{2+} or Mg^{2+} is also present in the cleft. Domain 1 and domain 3 are structurally related and probably evolved from gene duplication[31]. Domain 2 and domain 4 are considered as insertions between domain 1 and domain 3. Considering the direction of polymerization of monomers, G-actin is additionally separated into the (-)/pointed end (domain 2,4) and (+)/barbed (domain 1,3)[31]. Structural data has led to the hypothesis that G-actin can be present in two conformational states which is dependent on the bound nucleotide. Particularly the preference of actin binding proteins (ABPs) like cofilin and profiling towards ADP-G-actin and ATP-G-actin compared to fibrous actin (F-actin) enabled early theories of nucleotide-dependence of G-actin conformations[32].

Direct structural insights into the nature of F-actin have remained elusive in recent years[33]. Crystals for X-ray crystallography are hard to obtain as the natural polymerization/depolymerization generate filaments of varying lengths. Data obtained by electron-microscopy shows F-actin to be consisting of

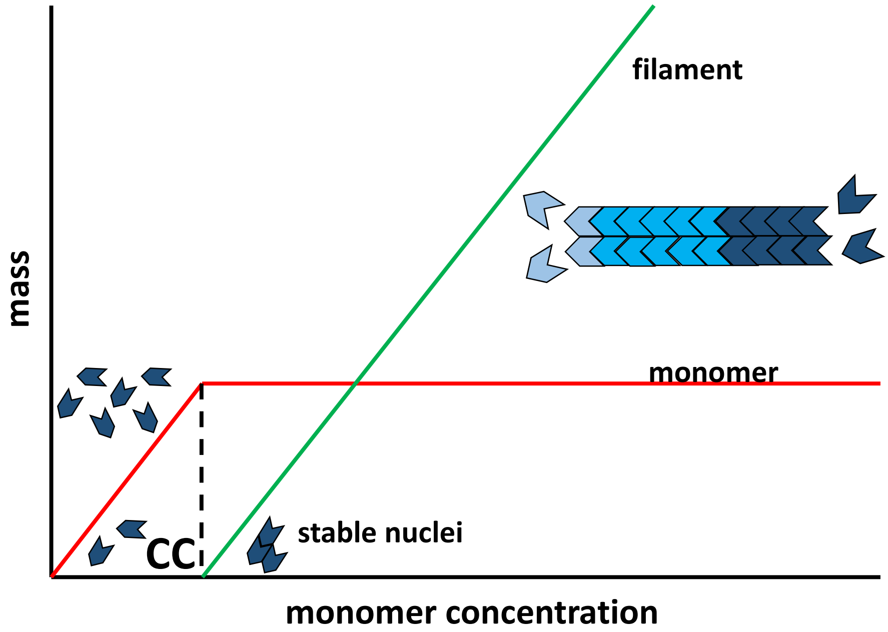


Figure 6: Schematic of the relationship of monomer concentration and formation of fibrous actin under conditions which enable polymerization. After the critical concentration (CC, dashed line) is reached additionally supplied actin monomers (red line) do not significantly increase the effective monomer concentration since all monomers are integrated into the filament upon polymerization which results in filament growth linearly correlated with supplied monomers (green line)[34].

two chains being intertwined in a right-handed "long-pitch" helix in which the inner domain is closer to the axis of the helix (Figure 4a)[35]. Today there are two models of the atomic nature of actin monomers in F-actin. The model proposed by Oda et al. in 2009 was obtained by optimizing the crystal structures of G-actin monomers into the helix of F-actin and then comparing it to X-ray fiber diffraction patterns. The model was improved by application of simulated annealing and other *in silico* techniques in order to improve the fit to the diffraction pattern. According to this model the transition from G-actin to F-actin involves a 20° rotation of the outer domain in relation to the inner domain at a right angle to the helix. This rotation reduces the twist between the inner and outer domain and flattens out the molecule. Additionally, the D-loop of subdomain 2 is relocated inwards[26]. The Holmes et al. 1990 model of F-actin was obtained in a similar fashion to the Oda model from an X-ray diffraction pattern into which monomeric actin coordinates were fit[36,37]. The prime difference to the Oda model is that the Holmes model does not provide for structural alterations of the actin monomer as the monomers are just stacked within the actin fiber[38].

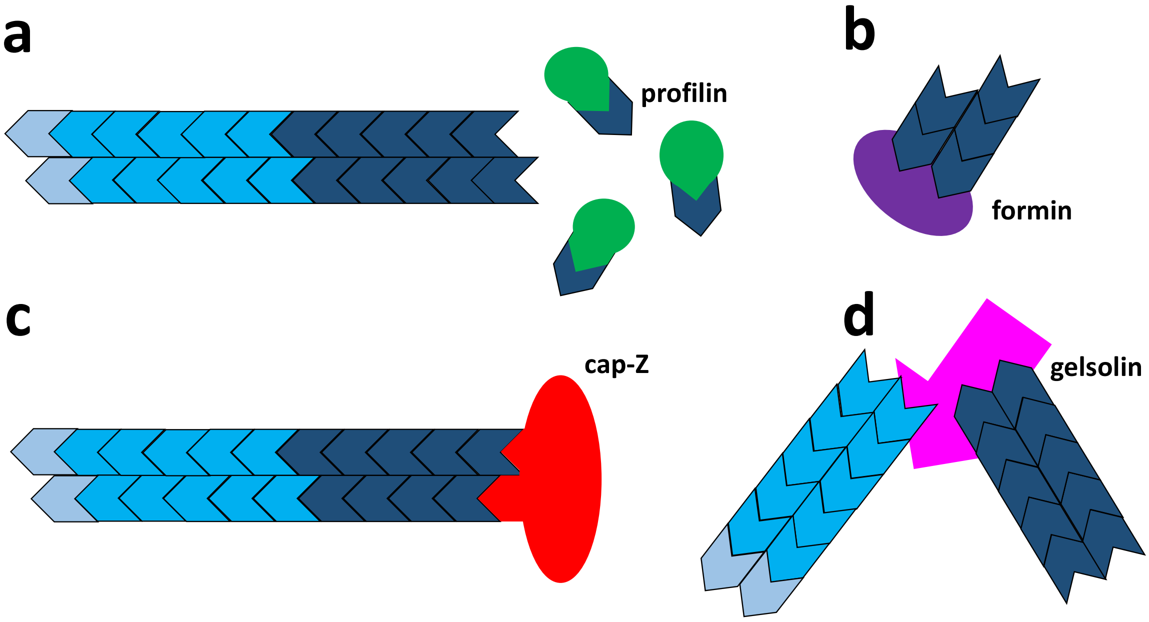


Figure 7: Simplified representations of actin binding proteins disrupting the fibril dynamics. **a)** Mechanism of monomer sequestration. Profilin is able to bind to G-actins to prevent association at the barbed end thus reducing the effective critical concentration and subsequently polymerization[39]. **b)** The FH2 domain of formins acts as a nucleation enhancing molecule by stabilizing loosely associated oligomers into steady nuclei[40]. **c)** Capping of actin filaments is mediated by molecules attaching at both ends and circumvent the addition of novel monomers. Capping is a function shared by many F-actin severing molecules like Gelsolin[41]. **d)** Severing of F-actin by Gelsolin occurs with near 100 % efficiency at any point of the filament after which Gelsolin caps the pointed end.

5.1.4 Actin polymerization and depolymerization

The polymerization of G-actin monomers into actin fibers is considered to happen in three stages. A nucleation, elongation, and a steady phase (Figure 5)[34]. Nucleation happens in regions of high-density ATP-G-actin when at least three monomeric subunits begin to form a stable nucleus. It is important to note that the formation of nuclei happens when a critical concentration of ATP-G-actin monomers is present (Figure 6). However, not all nuclei remain stable long enough for subsequent elongation to occur nucleation which leads to elongation is therefore heavily regulated by nucleation factors promoting nucleation and enabling stable trimers. In solution the assembly of nuclei is characterized by a lag phase which sees the typical formation of unstable oligomers. Eventually an oligomer will form a stable seed of three to four subunits. These stable nuclei then polymerize fast into a continuously growing actin fiber. It is important to realize that the filament grows in both directions with different rates. The barbed end (+) sees the addition of new sub unites 10 times faster than the pointed end (-).

The elongation process will then continue until the steady state phase which is characterized by a

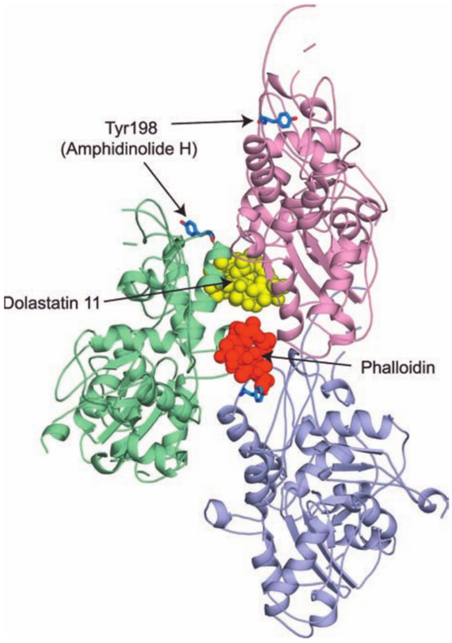


Figure 8: Schematic of experimentally determined binding sites of actin-stabilizing natural products phalloidin, dolastatin 11, and amphidinolide H. Phalloidin depicted in red spheres was identified to bind the interface between three different actin monomers (pink, green, purple cartoon)[42]. Dolastatin (yellow spheres) has been shown to bind the interface made up of domains 1 and 4 of two adjacent actin monomers[43]. Amphidinolide H (not depicted) binds covalently to the yeast actin residue Tyr198 (depicted in sticks) and its mammalian homologue Tyr200 (image taken from Allingham et al. 2006)[27,44].

continuous exchange of G-actin monomers at the end of the filaments without any detectable change in F-actin mass. Upon reaching the steady state phase the remaining G-actin monomers are represented by the so called critical concentration (CC). The critical concentration is representative for the ability of the remaining G-actin solution to polymerize. Under standard *in vitro* conditions the CC of G-actins is 0.1 µM above which the solution of G-actin will polymerize while F-actin will depolymerize if the concentration is lower. Experiments in which the polymerization/depolymerization was blocked at either (+) end or (-) end did show that the CC is different for both ends. While the aforementioned CC 0.1µM is measured primarily for the (+) end the (-) end has a CC of 0.8 µM. This difference leads to the observation of fiber "treadmilling"[45]. In the steady state phase there is constant fluctuation of the CC at both ends which eventually leads to the (+) end being constantly elongated while the (-) de-polymerizes. Under these conditions the effective length of the actin filament does not change while the fiber apparently moves into the direction of polymerization. Since the assembly of G-actin into the filament is dependent on the incorporation of ADP or ATP into the monomer. With increasing incorporation of ADP-actin into the filament the CC of the (+) end becomes identical to the CC of the (-)

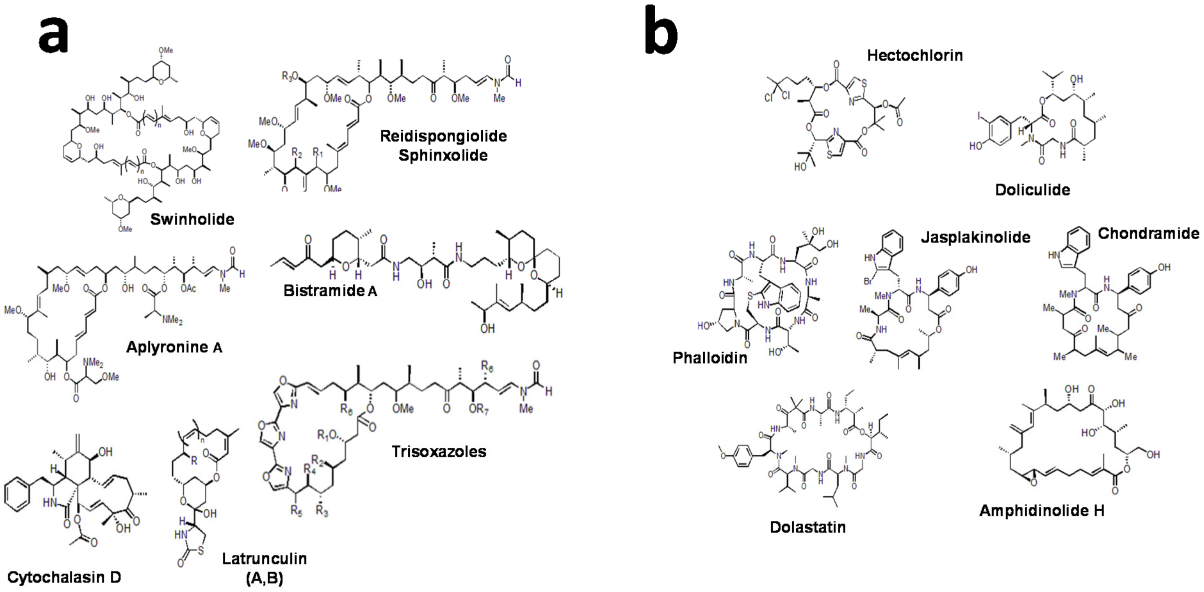


Figure 9: Overview for 2D depictions of natural product actin binders[27]. **a)** The two modes natural products engage in upon binding to actin is filament severing and over stabilization. With the exception of latrunculins, all known severing NPs bind to the barbed end while latrunculins are located in the nucleotide binding cleft. Most of the severing NPs share a common molecular layout with a large macrolide ring and a flexible tail. **b)** Natural products which artificially over stabilize the actin filament share a common macrolide ring but decorations not as large as the tail section of the severing NPs.

end. Similar to the polymerization process in general, the CC is also highly dependent on the environment. Polymerization dynamics are strongly influenced by pH, Ca^{2+}/Mg^{2+} concentration, ionic strength of the solution, as well as temperature. Depending on the concentration of Ca^{2+}/Mg^{2+}, the CC can be high despite a high concentration of G-actin monomers.

5.1.5 Disruption of actin dynamics

As the dynamics of actin polymerization/depolymerization can be influenced at each of the aforementioned steps, a multitude of proteins is translated to ensure controlled actin dynamics[1]. Nucleation factors containing WH2-domains such as spire, cordon-bleu, leiomodoin and VopF/VopL are considered to bind to unstable actin oligomers to form stable nuclei while proteins such as WASP/WAVE/Scar enable nucleation along existing filaments in order to achieve branching[46]. Additionally, formin proteins associate via their formin homology domain (FH1/FH2) at the barbed end of profilin-actin and enable continuous addition of new monomers to the filament[40]. The gelsolin family of actin binders including gelsolin, adseverin, villin capG, advillin, supervillin and flightless I

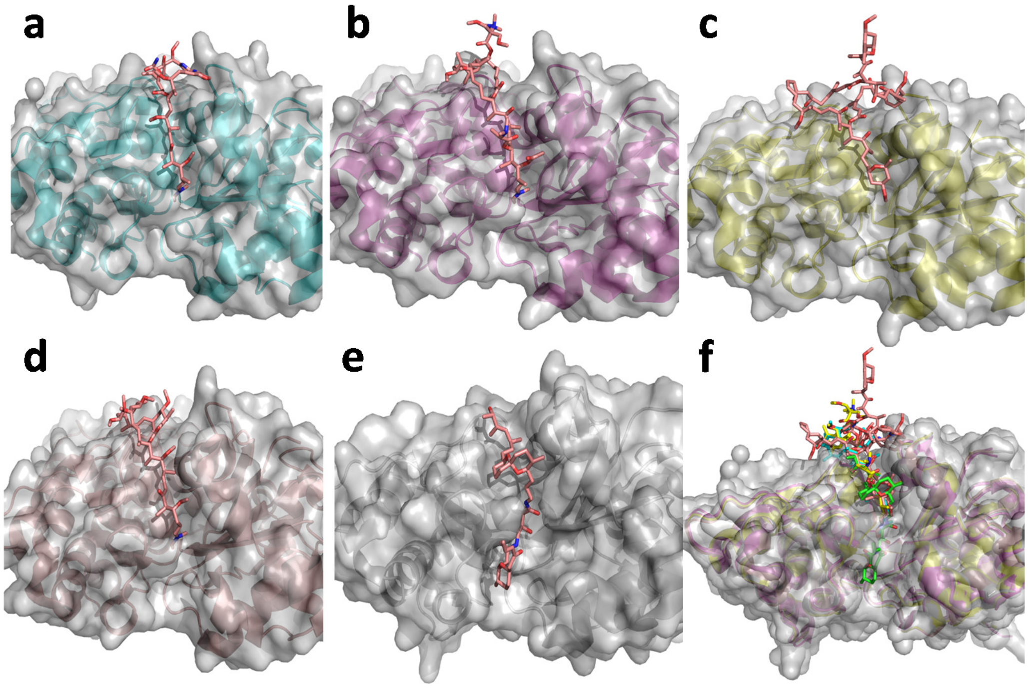


Figure 10: Example crystal structures of F-actin severing natural products. **a-e)** Kabiramide C, aplyronine, swinholide, reidispongiolide A, bistramide A all bind at the barbed extending from the periphery into the cleft between subdomains 1 and 3[47–51]. **f)** Superposition of all severing NPs. Due to a different overall structure in bistramide A compared to other severing NPs, it is the only one which extends further into the subdomain 1,3 cleft while the other NPs have the macrolide located on the periphery and the tail inserted into the barbed end.

serves both as nucleation promoters combined with capping properties at the barbed end (Figure 7)[41]. In addition, gelsolin binding to F-actin induces distortions to the filament which results in severing after which it also caps the barbed end. Contrary to enhancing nucleation, the cofilin family, consisting of ADF/cofilin, twinfilin, Abp1/Debrin, coactosin, and the glia maturation factor have been identified to enhance pointed end depolymerization, increased release of Pi from ADP-Pi-F-actin, and inhibit the binding of myosin, tropopmyosin and phalloidin[52,53]. The binding of proteins of the gelsolin family to F-actin is primarily controlled via pH and reduces the longitudinal interaction between subunits in the filament. Compared to gelsolin, profilin binding to ATP-G-actin greatly enhances nucleotide exchange at the barbed end, resulting in 1,000 fold increase of barbed end polymerization while profiling is prevented from association with the pointed end where polymerization effectively halts. Additional to proteins, large natural product (NP) molecules have been identified to associate in a variety of ways with actin and disrupt polymerization and depolymerization dynamics[27,48]. Several natural inhibitors of actin are derived from deep sea sponges, bacteria, and fungus which produce these molecules as toxins

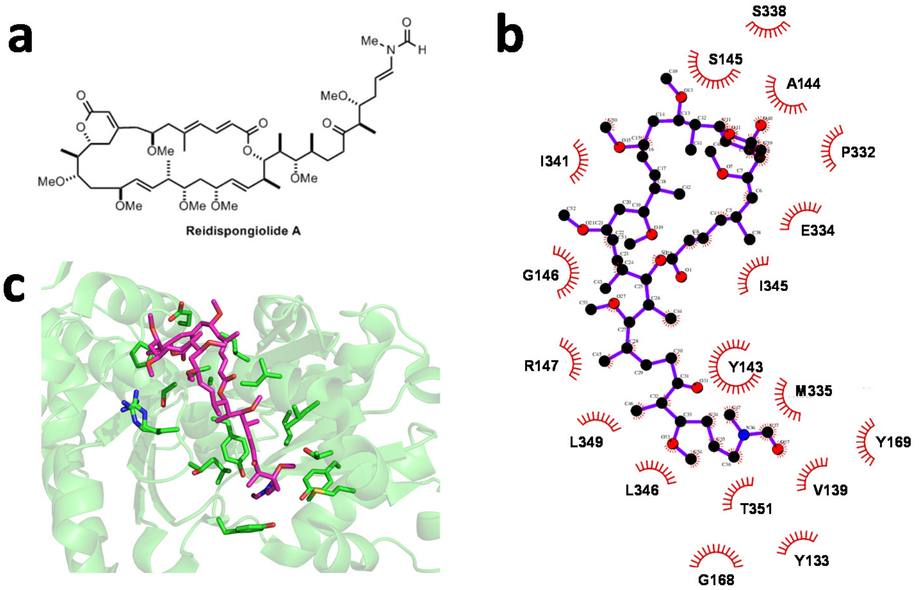


Figure 11: Overview of reidispongiolide A structural information. **a)** 2D depiction of reidispongiolide A generated with ChemDraw Pro 12 (CambridgeSoft, Cambridge, MA, United States). **b)** Ligplot+ depiction of interactions between redA and actin. The majority of interactions are van der Waals type and weak electrostatics. Hydrogen bond geometries indicating strong hydrophilic interaction were not satisfied[54]. **c)** Crystal structure showing the coordinates of redA interacting with actin.

for both disabling prey and fending off predators[27,55] The disruptive effects of these natural products are at the focus of research into understanding the dynamics of actin as well as promising precursors to drugs dealing with diseases characterized by defective actin dynamics. Natural products targeting actin are generally classified according to their site of interaction with actin and the effect on dynamics. There are two primary sites of interaction with actin, one being the ATP binding cleft which till today is targeted exclusively by latrunculin the other is the barbed end and its periphery[56–58]. The effects of natural products binding to actin is characterized by the two mechanisms of filament stabilization and inhibition of filament assembly or destabilization of the filament. Latrunculin A is the most prominent NP which binds to the ATP binding cleft and subsequently reduces binding cleft and subsequently reduces the inter-domain flexibility of the monomer which leads to impaired polymerization[59]. The main molecular feature by which this is accomplished is the thiazolidinone head group which interacts with Tyr69, Asp157, Arg183, Thr186, Arg206 and Arg210 via a distinct hydrogen bonding pattern[59]. The macrocycle provides bulk hydrophobic interaction with the monomer which then speculated to result in reduced degrees of freedom of the twist angle[59]. The effect of latrunculin interaction with actin

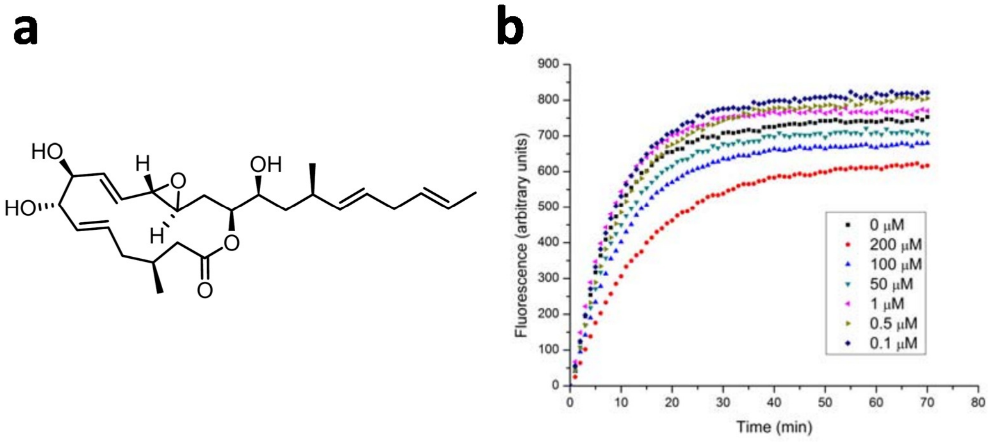


Figure 12: **a)** 2D structure of iriomoteolide 3a generated with ChemDraw Pro 12. **b)** Effects of different concentrations of iriomoteolide on actin polymerization. On concentrations below 1 μM iriomoteolide displays behavior indicative of polymerization enhancing proteins like nucleation promoters. On concentrations above 50 μM iriomoteolide, inhibition of actin polymerization was detected. Figure taken from Figure 33b of the Master's thesis of Andrea Unzue Lopez entitled *Iriomoteolides as novel chemical tools for the study of actin dynamics*, University of Zurich 2012.

on the polymerization dynamics is limited to sequestration of the monomers[56]. Contrary to latrunculins, other natural products identified to this day bind at the barbed end and the periphery such as trisoxazoles, reidispongiolides/sphinxolides, aplyronine A, scytophycin, swinholide A misakinolide A, lobophorolide, bistramide A, kabiramide C, cytochalasin, phalloidin, jaspakinolide, chondramide, doliculide, dolastasin, amphidinolide H and hecochlorin. These molecules engage in the two mentioned modes of disruption of filament dynamics. Kabiramide (trisoxazole), swinholide A, misakinolide, reidispongiolide, sphinxolide B, mycalolide (trsoxazole), aplyroninie, bistramide and cytochalasine all destabilize the formed actin filament while phalloidin, jasplakinolide, chondramide, doliculide, dolastatin, aphidinolide H and hectochlorin stabilize the actin filament (Figure 9,10)[27]. Most of the stabilizing NPs share a common general makeup, as in that they feature a large macro cycle made up of and decorated with hydroxyl, carbonyl, ether, methyl groups and a tail section (bistramide A and cytochalasin D are the two exemptions. Structural data provided insight into how these NPs probably destabilize the filament by association of the macrocycle with the hydrophobic patch located at subdomain 3 and subsequent insertion of the tail section into the barbed end (subdomain 1, 3)[48]. Destabilization of the filaments is primarily achieved via sequestering monomers from forming nuclei as well as association to the filament or severing an existing F-actin strand. While sequestration of monomers is mainly done by all destabilizers, severing and capping properties are not equally distributed. Natural products which stabilize the filament are structurally less understood as X-ray crystallography has not been able to provide such structures (Figure 8). Enhanced polymerization in their presence makes crystallization of complexes nearly impossible which results in the majority of information being derived from x-ray fiber diffraction data, electron microscopy and biochemical assays. A cyclic depsipeptide is the most common feature of stabilizers present in phallodin, jasplakinolide, chondramide doliculide, dolastatin but not in amphidinolide H and hectochlorin. The mode of interaction of stabilizers understood through two examples. Phalloidin bridges three monomers in F-actin which appears to be stabilizing two long-pitch strands via interaction with residues 198-201 of filament subunit n, residues 73-75 of subunit n+1, and residue 179 of a 3rd diagonal subunit[42]. Similarly, dolastatin 11 also bridges two long-pitch strands but by providing the strand interface at subdomain 4 of one monomer with subdomain 1 of the diagonal subunit[43].

Iriomoteolide 3a has been previously identified as a cytotoxic natural product of marine origin and its total synthesis was achieved in 2009[60–63]. It is a 15-membered macrolide with an allyl-epoxide which was isolated from *Amphidinium sp.* (HYA024) and characterized by 2D-NMR. The cytotoxicity was established for human B lymphocyte DG-75 (IC50 of 0.08 µg/ml) and Raji cells infected with Epstein-Barr (IC50: 0.05 µg/mL). Previous studies on the effect of iriomoteolide 3a on actin polymerization show that on low concentrations iriomoteolide displays increased polymerization similar to molecules which promote nucleation and polymerization[64]. On high concentrations (50 µM, 100 µM, 200 µM) the results show that iriomoteolide exhibits a polymerization inhibiting property (Figure 12b).These results as well as the similar makeup of iriomoteolide and natural products like reidispongiolide A (Figure 11a, 12a) being a macrolide and tail was what initiated the investigations of iriomoteolide 3a binding towards actin as presented in this thesis chapter.

5.2 Methods

5.2.1 Energy minimization and simulation protocols

Coordinates for monomeric G-actin bound to gelsolin, ATP, and calcium (CA) were downloaded from the protein databank (PDB-ID: 1EQY)[41]. Coordinates for reidispongiolide A bound to G-actin were downloaded from the protein data bank (PDB-ID: 2ASM)[48]. Gelsolin, waters, and co-solvent molecules were manually removed from 1EQY and 2ASM. Any missing atoms from 1EQY G-actin residues ranging from 5 to 375 were added via CHARMM[65–67]. Parameters for all amino acids were created with the *pdb2gmx* module of GROMACS (version 3.3.6) for the CHARMM 27 protein force field parameters[68–70]. Parameters for calcium and ATP where also taken from the CHARMM 27 force field. For both systems (actin, ATP, CA, and iriomoteolide/reidispongiolide A) threonine 6 was capped with an n-terminal acetyl cap while the C-terminal phenyl alanine 375 was charged negatively[71]. The 3-dimensional structure of iriomoteolide 3a was generated via Pymol and a TRIPOS MOL2 file was then generated manually[19,72]. Reidispongiolide A coordinates were taken from PDB-ID 2ASM and then converted into TRIPOS MOL2 format by hand. CHARMM General Force Field (CGenFF v. 2b7) parameters for iriomoteolide 3a and reidispongiolide A were obtained via uploading the respective MOL2 files to paramchem.org[73–76]. To make CGenFF parameter stream files (.str) compatible with simulations run in GROMACS (.itp), bond, angle, dihedral, improper units were converted from kcal/mol to kJ/mol and Ångstrom to nanometers respectively. All 1-4 interactions were enumerated and pre-calculated via unit conversion of the original CHARMM parameters[77].

The subsequent preparation work flow is identical for all simulations of redA (bound and unbound) as well as iriomoteolide 3a (10 copies unbound and 1 copy unbound). Deploying the GROMACS package *editconf*, a dodecahedric box around the protein was created with a periodic boundary 1.2 nm distanced from the most outlying atoms of G-actin. For the simulations containing 10 copies of iriomoteolide simultaneously, the box was then filled with 10 copies of iriomoteolide at random positions determined by the *genbox* module of GROMACS. For the single copy iriomoteolide 3a/redA simulations, irio3a and redA were placed at random positions of the box. The system was then filled with repetitive units of pre-equilibrated SPC water coordinates parameterized for the TIP3P water model[78–80]. To neutralize the system sodium and chlorine ions were added at a concentration of 150 mM via the *genion* module of GROMACS.

Each system was then minimized for 10,000 steps via the steepest descent minimizer or prematurely stopped if the total energy of the system was lower than 10 kcal. A minimization step size of 0.01 kcal/mol was selected. After energy minimization, the system was equilibrated in a simulation for 1 ns with random velocities and at constant temperature, pressure and number of particles (NPT) in which restraints were applied to all heavy atoms of G-actin, iriomoteolide, reidispongiolide A, ATP and CA while the and NaCl ions and water could be propagated unrestrained. All production runs were started with new initial, random velocities.

All simulations (including equilibration) were performed with a velocity rescaling thermostat and a Parinello-Rahman barostat at 1 atmospheric pressure unit and a temperature of 310 K[81–83]. The time step of the simulation was set to 2 fs. Long range vdW and electrostatic interactions were calculated with simple cut-off scheme and a particle mesh Ewald (PME) cut-off scheme at 1.2 nm[84,85]. The LINCS algorithm implemented in GROMACS was applied to constrain all atoms in the system[86]. For the 10-copy iriomoteolide simulations 10 separate simulations were run for 100 ns with different initial velocities but identical positions of the respective iriomoteolides within box.

For the 25 and 5 single iriomoteolide simulations, the simulation times were 100 ns and 700 ns respectively. The positions of iriomoteolide in the 25 100 ns simulations are identical to the 5 700 ns simulations each started 5 times with random velocities. Out of 25 100 ns iriomoteolide simulations, 5 simulations were prolonged to a total of 300 ns each.

Five simulations of bound reidispongiolide A were run for 100 ns each while the 5 simulations of unbound redA were simulated for 600 ns. All simulations were started with random initial velocities. Each simulation was performed on 64 Xeon 5560 processors clocked at 2.8 GHz each via MPI for GROMACS 3.3.6 on 64 cores per simulation. A reference .mdp file which was the template for all production run simulations can be found in supplementary material.

5.2.2 Distance evaluation and contact frequencies

Prior to analyzing the trajectories, all waters and ions were removed so that only G-actin, ATP and iriomoteolide/reidispongiolide A were contained in within the system. Movement of atoms across the PBC was removed in three steps via the *trjconv* GROMACS module. First, single atomic movement over the PBC was removed so that if a single atom would cross the PBC the entire molecule would be propagated (*-pbc mol*). In a second *trjconv* step G-Actin, ATP, and iriomoteolide/reidispongiolide A were clustered (*-pbc cluster*) in the respective frame. Finally, only the actin protein was centered in the box and all ATP, irio3a/redA movement was to happen in respect to the centered protein (*-center*). Finally, all trajectories were concatenated into a single trajectory via *trjcat -cat* module implemented in GROMACS.

The distances of iriomoteolide with every residue of actin as well as other iriomoteolide molecules in the same box were calculated via *g_dist* in GROMACS. For iriomoteolide, reidispongiolide A, and the respective residue, the centers of mass (COM) were considered for evaluation of distances. Contact frequencies for each residue were calculated by evaluating if the center of mass distances between irio3a/redA and each residue was equal and lower than 1 nm. If the threshold was crossed each snapshot would be registered with a binary value of 1. For the simulations containing 10 iriomoteolides, all irio3a copies were considered per residue. If the threshold was not crossed the snapshot would be registered with 0. The relative contact frequency (RCF) for each residue for the given iriomoteolide simulations (5x700 ns, single iriomoteolide; 25x100 ns, single iriomoteolide; 10x100ns ten iriomoteolides) was then calculated by:

$$C_i = \frac{N_i}{max\left(N_{1..n}\right)} \tag{9}$$

where **C$_i$** is the contact frequency for residue **i**. **N$_i$** for each residues **i** is the number of snapshots for which the contact binary was set to 1. **max(N1..n)** is the largest **N$_i$**. All contact frequencies are thus annotated relative to the contact frequency of the "most visited" residue which has a RCF of 100%. Heat maps on the protein surfaces were then generated via Pymol by substituting the b-factor column with the relative contact frequencies and then applying a green/red color scheme where red is high contact frequency and green is low contact frequency[19].

5.2.3 Binding mode analysis: clustering and cut-based free energy profiles

Prior to the cluster analysis and cut-based free energy profile (cFEP) generation all alpha carbons of G-actin were aligned to the starting structure of the respective production runs for every snapshot[87,88]. Cluster analysis was performed using tree-based clustering included in CAMPARI, taking into accounts all 700 ns simulations for which iriomoteolide traversed the system to the barbed end of G-actin[89,90]. RMSD-based clustering was performed with threshold radius (*CRADIUS*) of 2.5 Å, a tree height (*BIRCHHEIGHT*) of 9 and a coarsest threshold (*CMAXRAD*) of 14 Å. This procedure was repeated for the combined simulations including the original 3 x 700 ns simulations plus 5 x 300 ns simulations which were elongated from the 25 x 100 ns simulations. The clustering was performed on all heavy atoms of iriomoteolide. Subsequent to clustering a cut-based free energy profile (cFEP) was generated in WORDOM (version 0.22)[91,92].

The free energy profile was generated by deploying a method based on the equilibrium kinetic network which preserves the free energy barriers[87,88,93]. This methods emulates the cuts in flow-networks, and the computed profile is therefore named cut-based free energy profile. The nodes and links of the equilibrium kinetic network are the clusters which were generated via the tree-based clustering as outlined above and the direct transitions between them sampled along the MD runs, respectively. For each node, all nodes are partitioned into two groups A and B by applying the mean first passage time (MFPT) to the reference node as an order parameter. The free energy is related to the maximum flow between sets A and B and calculated as $G= -kT \ln(Z_{AB}/Z)$, where Z_{AB}/Z is the relative partition function which represents the statistical weight of the transitions between sets A and B. The result is a one-dimensional profile along the reaction coordinate Z_A/Z (i.e., the relative partition function representing the statistical weight of set A) which preserves the barrier height between the free energy basins.

5.3 Results

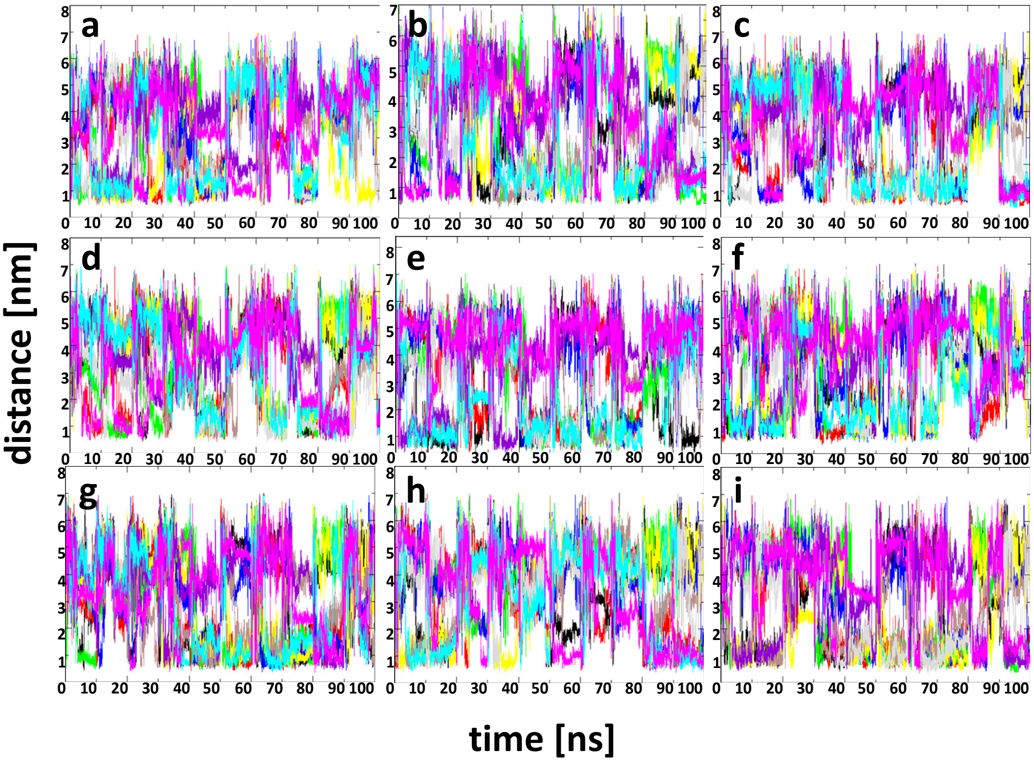


Figure 13: Overview of inter-iriomoteolide distances for the 10x100 ns simulations with an irio3a : actin ratio of 10 : 1. **a-i)** Each color (purple, cyan, green, blue, black, red, brown, yellow, gray, magenta) shows the distance of the first iriomoteolide in the system with the remaining nine. Distance evaluations of other iriomoteolides were omitted as aggregation was considered severe already. Over the course of all simulations there is severe aggregation denoted by distances around 0.5 nm for more than 70 ns for several molecules. Center of mass distances were generated via *g_dist* implemented in GROMACS (v 3.3.6) and plotted in GRACE (5.1.23, Evgeny Stambulchik)[69].

5.3.1 Simulations of high concentration iriomoteolide result in aggregation.

Due to large size of monomeric actin and the resultant particles in the simulation exceeding 100,000, speeding up binding simulations was considered critical. The likelihood of interaction of a single iriomoteolide with actin is higher with increasing concentration of the ligand molecule. The box size chosen for the simulation and the 10 copies of iriomoteolide in the system resulted in a concentration of roughly 16 µM iriomoteolide. For simulations containing 10 simultaneous iriomoteolides with G-actin,

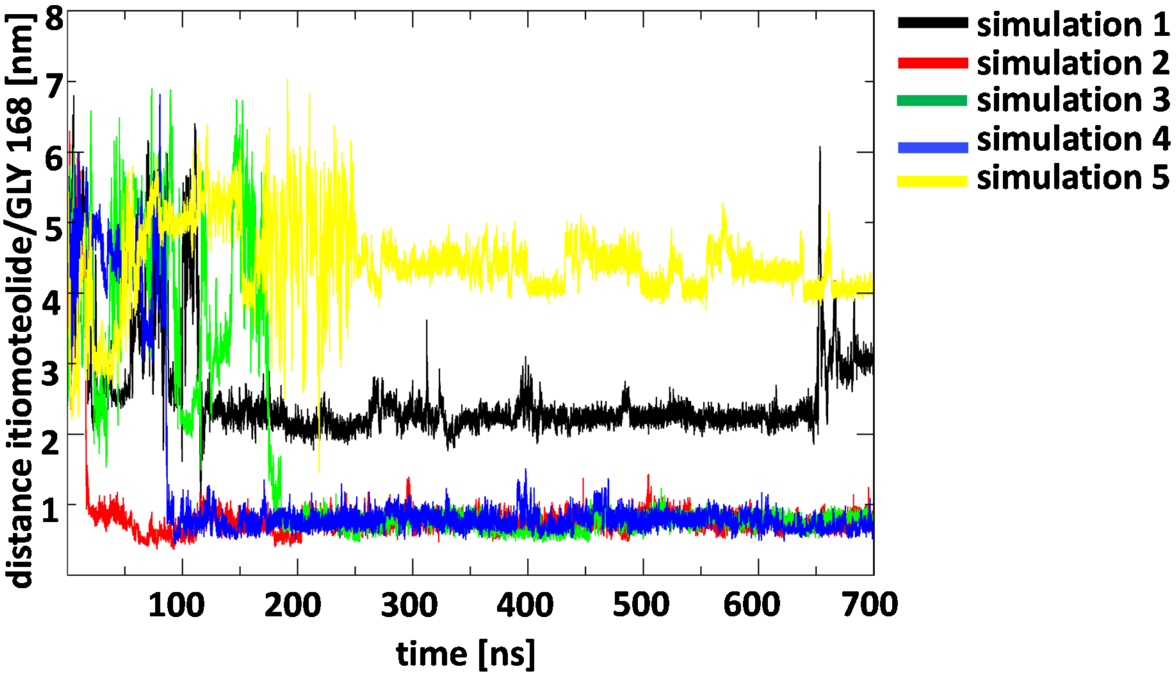


Figure 14: Distances of single iriomoteolides with Gly168 for the 5 x 700 ns simulations with an irio3a : actin ratio of 1 : 1. Simulations 2-4 show center of mass distances between irio3a and Gly168 smaller than 1 nm after 20 ns, 90 ns, and 170 ns respectively. Simulations 1 and 5 bind with a 2nd and 3rd interaction site after 120 ns and 250 ns. Distances plotted in GRACE (5.1.23, Evgeny Stambulchik).

a simple distance evaluation from one copy to the other was carried out. The analysis shows that at least two iriomoteolide were aggregated at some point in all simulations (Figure 13 a-i). The aggregates were characterized by the individual molecules being within a center of mass (COM) distance of 1 nm. It is important to point out that due to its size and the center of mass of iriomoteolide is roughly 0.5 nm from its outmost atoms. The shown aggregates remained on average for 10 ns before one iriomoteolide moved away and associated with different iriomoteolide molecules or the same cluster would be formed again.

5.3.2 High-concentration iriomoteolide simulations show that there are three hotspots of association with the actin monomer.

The primary site of interaction is around residue Met227 (Figure 15c, Supplementary table 1) with the largest relative contact frequency of 100 %. This hot spot of interaction with iriomoteolide extends up to 5 adjacent residues (upstream and downstream). A secondary and tertiary site of association were observed at residues Met283 and Gly168 respectively. High relative contact frequency values for

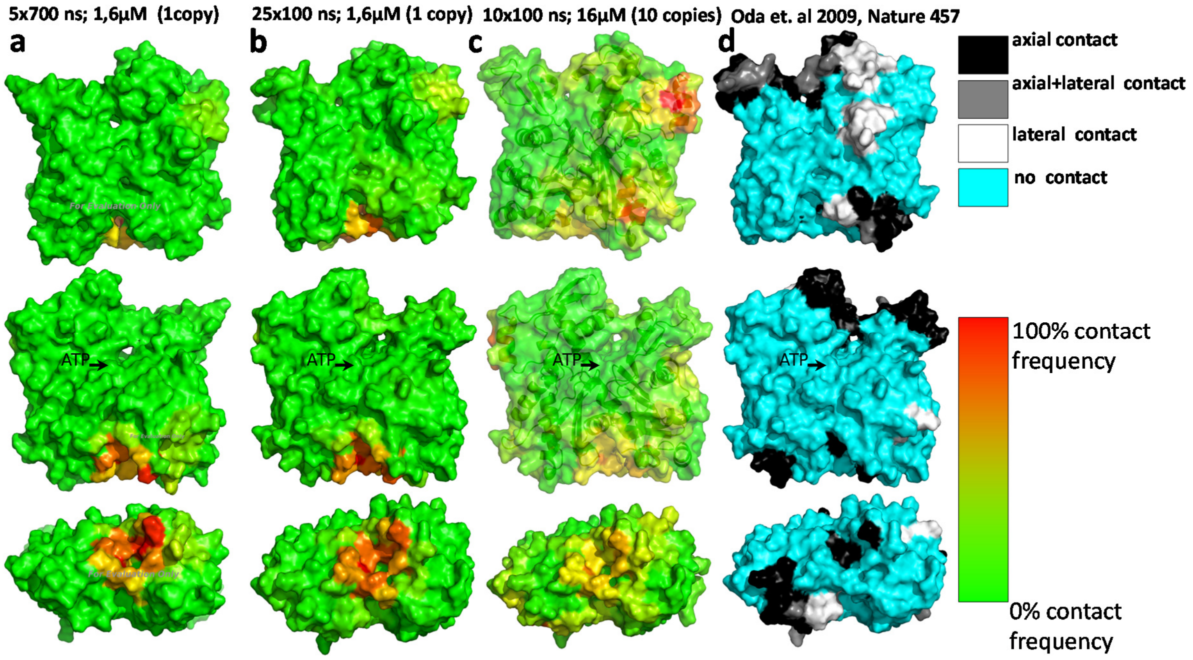


Figure 15: Summary view of the relative contact frequencies for all actin iriomoteolide simulation schemes. **a-c)** Heat map of relative contact frequencies for the three distinct simulation schemes. Top and middle panel represent front and back of G-actin, the bottom panel shows a top-down view of the barbed end. a) and b) show high similarity in the distribution and intensity of the respective high contact residues, primarily in the barbed end. c) shows lower C_i values for the barbed end and higher values for Met355. **d)** Annotation of axial/lateral contacts in F-actin with the adjacent monomers. The pointed end and the barbed end of the next monomer are interfaces where obstruction would lead to an effective sequestering of monomers.

sequentially adjacent residues attributed to the same location on the protein. The fourth peak in contact frequency (64 %) around Phe352 is closely located at the barbed end near Gly168. However, given the observed aggregation of iriomoteolide molecules, these results were considered to be severely impacted by the aggregates. Aggregates interacting with actin would restrict the single-molecule interaction as the iriomoteolide closest to the actin would be influenced by the surrounding iriomoteolides. To evaluate iriomoteolide binding behavior without aggregates, single-iriomoteolide simulations of different lengths were subsequently run.

5.3.3 Single-iriomoteolide simulations indicate preferred interactions with the barbed end of monomeric actin.

To analyze interactions of iriomoteolide with G-Actin without eventual effects of aggregation a 1:1 ratio simulation of iriomoteolide 3a and actin was chosen. A total of 30 simulations were conducted out of which 5 simulations were run for 700 ns, 25 simulations for 100 ns of which 5 were extended to 300 ns resulting in a cumulative simulation time of 7 µs. The 5x700 ns simulations initially run, were evaluated for the distance of iriomoteolide with Gly168, a central residue of the barbed end. The distance analysis showed that out of the five simulations with different positions of iriomoteolide within the box, three (Figure 14) show binding of irio3a to the barbed end of actin within 20 ns, 90 ns, and 170 ns respectively. All three irio3a molecules started from a distance to the barbed end which exceeded 4 nm. Subsequently, all three iriomoteolide molecules remained close to Gly168 for the remaining simulation time with distances fluctuating around 1 nm. Simulation 1 (Figure 14 black), for which no association to the barbed end was observed, indicated that for the first 110 ns of the simulation the iriomoteolide traverses the system with no fixed residence. Indicative for unobstructed movement through the system is the high fluctuation of distances within a short period of time. Following the 110 ns mark, the iriomoteolide remained at ~ 2.5 nm distance of Gly168 and stayed in place in relative stability until 650 ns. After fast dissociation, irio3a stably re-associated with the protein at a relative distance of 3 nm for the remainder of the simulation. In simulation 5 (Figure 14 yellow) the iriomoteolide moved away from Gly168 within the first 100 ns. Afterwards it showed high fluctuation of distances indicating repetitive movement over the periodic boundary of the system box. Following the 250 ns mark, this irio3a associated with actin and remains stable until 475 ns after which it entered a period of binding and rebinding events indicated by fast transitions between distances between 4.0 and 4.5 nm. This behavior continued until the end of the simulation indicating a site of weaker binding than for the barbed end in simulations 2-4 and simulation 1 (Figure 14 red, green, blue).

To further evaluate all residues for possible interactions with iriomoteolide a heat map displaying all contact frequencies was generated for the 5x700 ns and 25x100 ns simulations (Figure 15 a,b; Supplementary Table 2,3). The relative contact frequencies show a significant hot spot for interaction at the barbed end with Phe352 (C_i of 100%) representing the residue for which most of the interaction with iriomoteolide was observed. Secondary hot spots indicative for preferred interaction, albeit very weak in comparison, were located at Met227 and Met283. These sites of interaction are identical to the simulations with an iriomoteolide-to-actin-ratio of 10:1 however they differ in magnitude of the relative contact frequencies. For the 1:1 ratio simulation, the barbed end was the actin region of highest contact frequency while in the 10:1 ratio simulation this hotspot had the lowest relative contact frequencies (Figure 15a,c). Sites Met227 and Met283 displayed comparatively lower contact

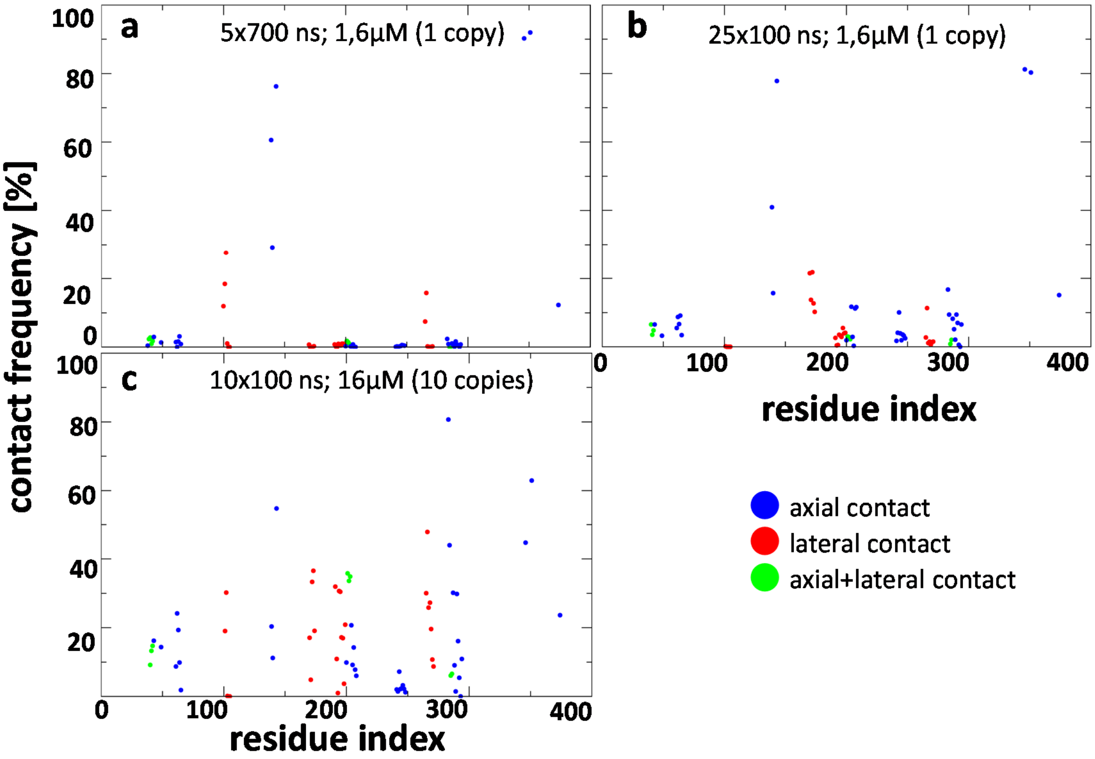


Figure 16: Summary of relative contact frequencies per residue of the individual simulation schemes, annotated only for residues identified to be in contact with adjacent monomers according to the Oda et al. 2009 model[26]. Blue dots show axial contacts (Figure 15d, black), red indicates lateral contacts (Figure 15d, white), and green indicates both axial and lateral contacts (Figure 15d, gray). **a)** Results for the 5 x 700 ns simulation show the highest contact frequencies (above 70 %) for residues involved in axial contacts. **b)** C_i profile for 25 x 100 ns simulations. Both single-iriomoteolide simulations schemes show very similar results indicating sufficient sampling. **c)** The C_i for the 10 copy iriomoteolide simulations show higher average contact frequencies for more residues which is likely caused by prolonged residue times per any of the iriomoteolide upon aggregation. The residues with the highest relative contact frequency are not shown however as Met283 and Met227 are not involved in inter-actin-interaction in F-actin.

frequencies in the 1:1 ratio simulations compared to the 10:1 simulations. Both simulation schemes showed very low contact frequencies outside of the observed hot spots. Overall the results indicated that the system was sufficiently equilibrated and that the barbed end represented the site of preferred interaction of iriomoteolide with actin.

5.3.4 Iriomoteolide at the barbed end would obstruct the interface of axial assembly of actin monomers.

The barbed end of monomeric actin is the interface for axial and lateral association in in the nucleation

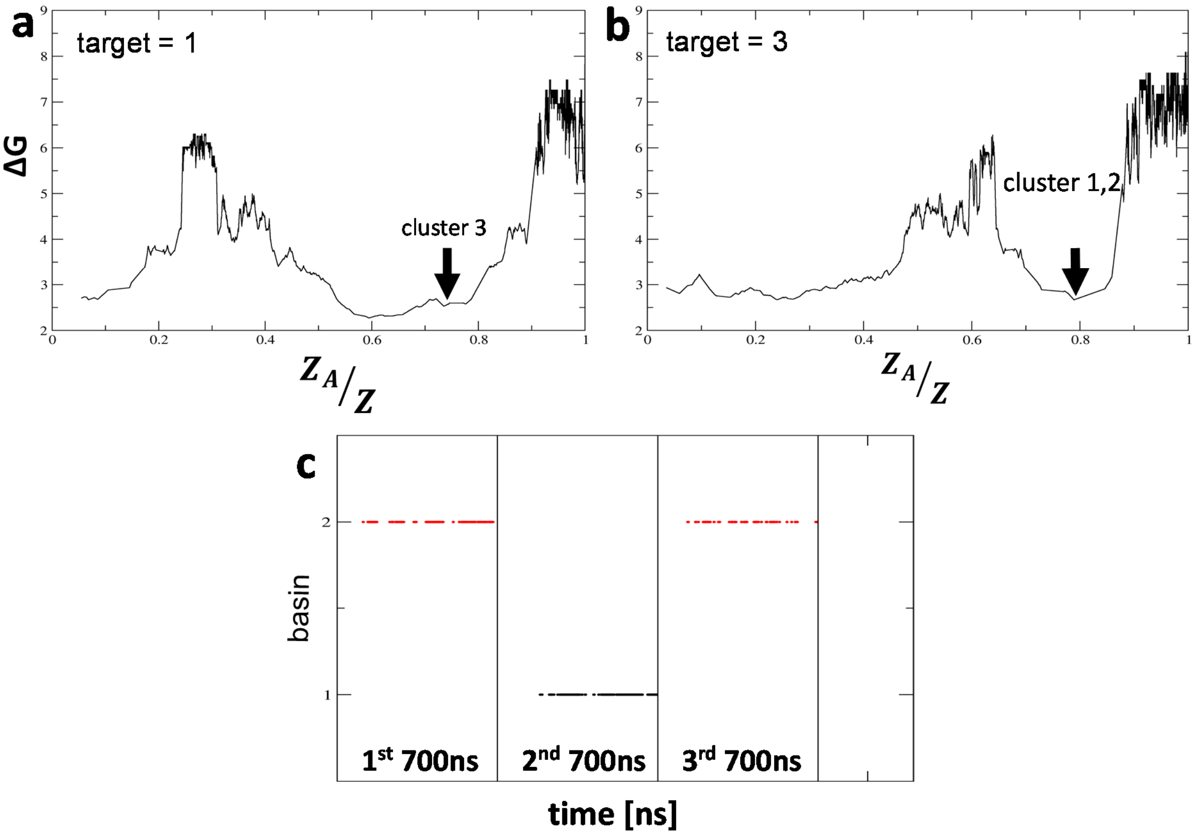


Figure 17: Cut-based free energy profiles and sampling frequency of the two major free energy basins. **a)** cFEP for the with the largest cluster being the target reveals two basins which are separated by a barrier of 6.4 kcal/mol. Cluster 3 is located within the second basin. **b)** cFEP with cluster 3 as the target. Similarly, this cFEP shows two free energy basins separated by a 6.5 kcal/mol energy barrier. **c)** Basin progression over all 5 x 700 ns. Although basin 2 was sampled in two out of the three simulations where iriomoteolide bound to the barbed end basin 1 was sampled more often within the simulation.

and polymerization of actin. It is an interface for a total of three actin monomers, n, n+1, n+2. A comparison of all heat maps with residues defined as being in contact with other actin monomers in F-actin (Figure 15 c,d) showed that out of the three hotspots identified the barbed end hotspot is the only one which would primarily interfere with association of G-actin subunits into nuclei or actin fibers[26]. This is further supplemented other natural productions being shown to bind to the barbed end, displaying inhibition of actin polymerization/nucleation[27]. Both lateral and axial interaction of one monomer with adjacent monomers would be obstructed by an iriomoteolide molecule binding inside the barbed end with high affinity. The second site of interaction, around Met283, is located outside the barbed end in subdomain 1 which features axial contacts with other actin monomers. Interaction site Met227 is on the periphery of several lateral contacts between actin monomers but featured no direct

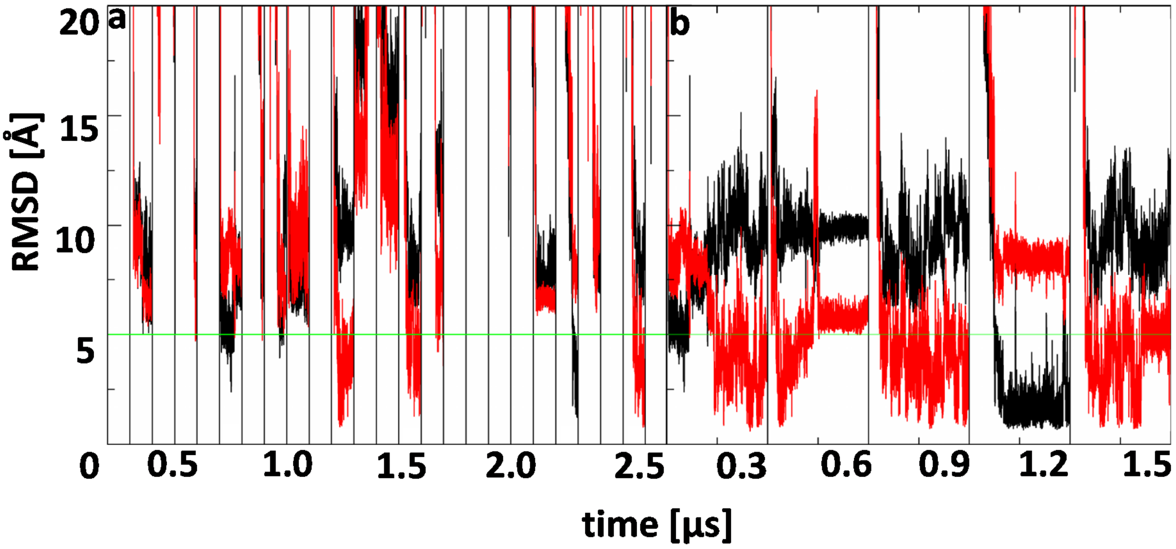


Figure 18: Time series of RMSDs of basin representatives from the 5 x 700 ns simulations for the 25 x 100 ns simulations. **a)** The analysis showed that in 25 simulations conformations close to the basin representatives (threshold of 5 Å, green line) were simulated five times. Within these five simulations basin 1 (black) and 2 (red) were sampled two and three times respectively. **b)** Extension of the five 100 ns simulations to 300 ns simulations revealed that three simulations kept sampling the basin for the entirety of the simulation while one simulation (1st column) exchanged the basin from 1 to 2. One simulation (2nd column) kept sampling iriomoteolide conformations slightly outside the threshold chosen but was none the less included in the later binding mode evaluation.

interaction with either. Thus, this hotspot was considered the least likely to interfere with nucleation or polymerization of actin given the Oda et al. 2009 model[26]. An analysis for residues only involved in inter-actin interaction showed high similarities between 25x100 ns simulations and 5x700 ns simulations (Figure 17a,b). Residues Leu346, Thr351 and Gly168 represent the residues with the highest contact frequencies above 70 %. All of these residues are located in the barbed end. Compared to the 10x100 ns simulations with 10 copies iriomoteolide those residues in the barbed end displayed a higher contact frequency. As previously shown the 10x100 ns simulations featured relatively higher contact frequencies for any residues including the ones interacting with n+1 and n+2 actin (Figure 16c). Additionally, this data showed clearly that the global maximum Met227 would not feature interference with actin nucleation/polymerization.

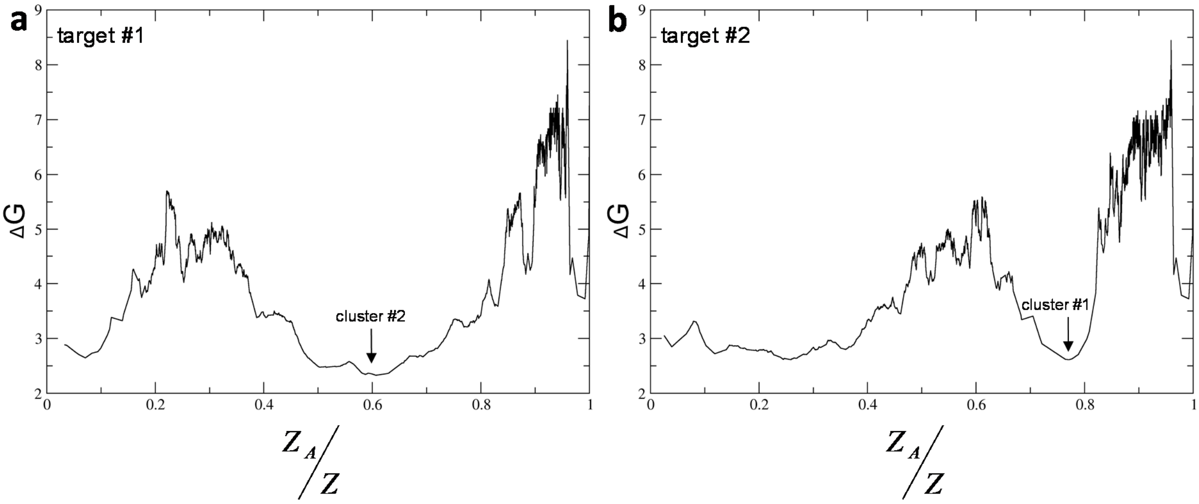


Figure 19: Cut-based free energy profiles for the supplemented simulations. **a,b)** Both cFEP plots show two basins separated between a 5.5 kcal/mol barrier. However the two largest clusters are now distributed distinctly between the basins.

5.3.5 Cluster- and cut-based free energy profile analysis revealed two distinct basins.

In order to identify possible binding modes for three 700 ns simulations in which iriomoteolide moved to the barbed end, a cluster analysis was conducted (Supplementary table 4). Deploying the tree-based clustering algorithm (developed by Vitalis et al. 2012 and implemented in CAMPARI) with a subsequent cut-based free energy profile analysis via the mean first passage time (MFPT) method, two free energy basins were identified (Figure 17 a,b). These basins are separated by an energy barrier 6.5 kcal/mol. Several small minima are located along this barrier but were neglected for further analysis. The three largest clusters of the clustering are located in these two basins. Interestingly, the two largest cluster (cluster 1: 11634 members, cluster 2: 8414 members) were located in the same basin. Behavior like this is indicative of a small kinetic distance between the respective clusters based the conformations sampled. It was projected that if the clustering cut-off (2.5 Å) was increased, clusters 1 and 2 would become a single cluster. The 3rd largest cluster is located in the second basin separated by the free energy barrier (7,444 members). To investigate if transitions between the two energy basins occurred, cluster progression analysis of clusters 1 (basin 1) and cluster 3 (basin 2) for the three simulations where iriomoteolide bound to the barbed end was generated. Basin 2 was sampled in two out of three simulations while basin 1 was sampled in a single simulation thus, no transition between basins within the respective simulation was recorded (Figure 17c). This was deemed indicative of the

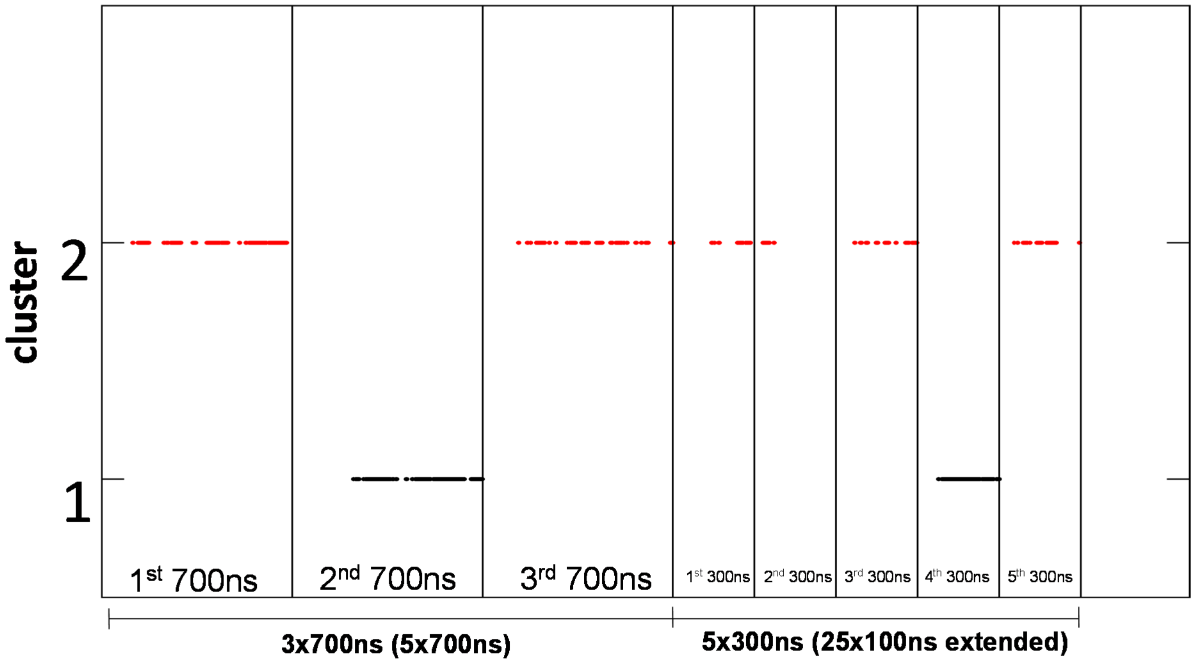


Figure 20: Cluster progression for the supplemented simulations. The results show that cluster 2 was simulated in 6 simulations while cluster 1 was simulated in 2 simulations. However, overall cluster 1 was sampled more often within the respective simulations indicating that this binding mode, once sampled, was more stable.

simulation data not being sufficient enough to derive exact binding pose prediction. To supplement the simulation data, RMSD evaluation for the representative clusters of the basins were performed for all 25x100 ns simulations. For five simulations a threshold of 0.5 nm was crossed showing that structures closely resembling the respective clusters were sampled (Figure 18 a). Three simulations sampled conformers closely similar to the representative of basin 2 with a RMSD being as low as 2.5 Å. In three simulations iriomoteolide conformations highly similar to representative of basin 1 were sampled out of which two were taken for extended sampling (Figure 18b). All chosen simulations were then extended to 300 ns at the end of the study. All extended simulations showed RMSD values below the threshold of 5 Å and where thus incorporated into the second generation of binding mode determination.

5.3.6 Supplemented simulations show two binding modes at the barbed end.

The evaluation of the tree-based clustering and subsequent cFEP for the combined simulation lead to a similar picture as the initial 3x700 ns simulations. Two distinct basins were identified separated by a barrier with 5.5 kcal/mol and several local minima in between. The barrier height was independent of

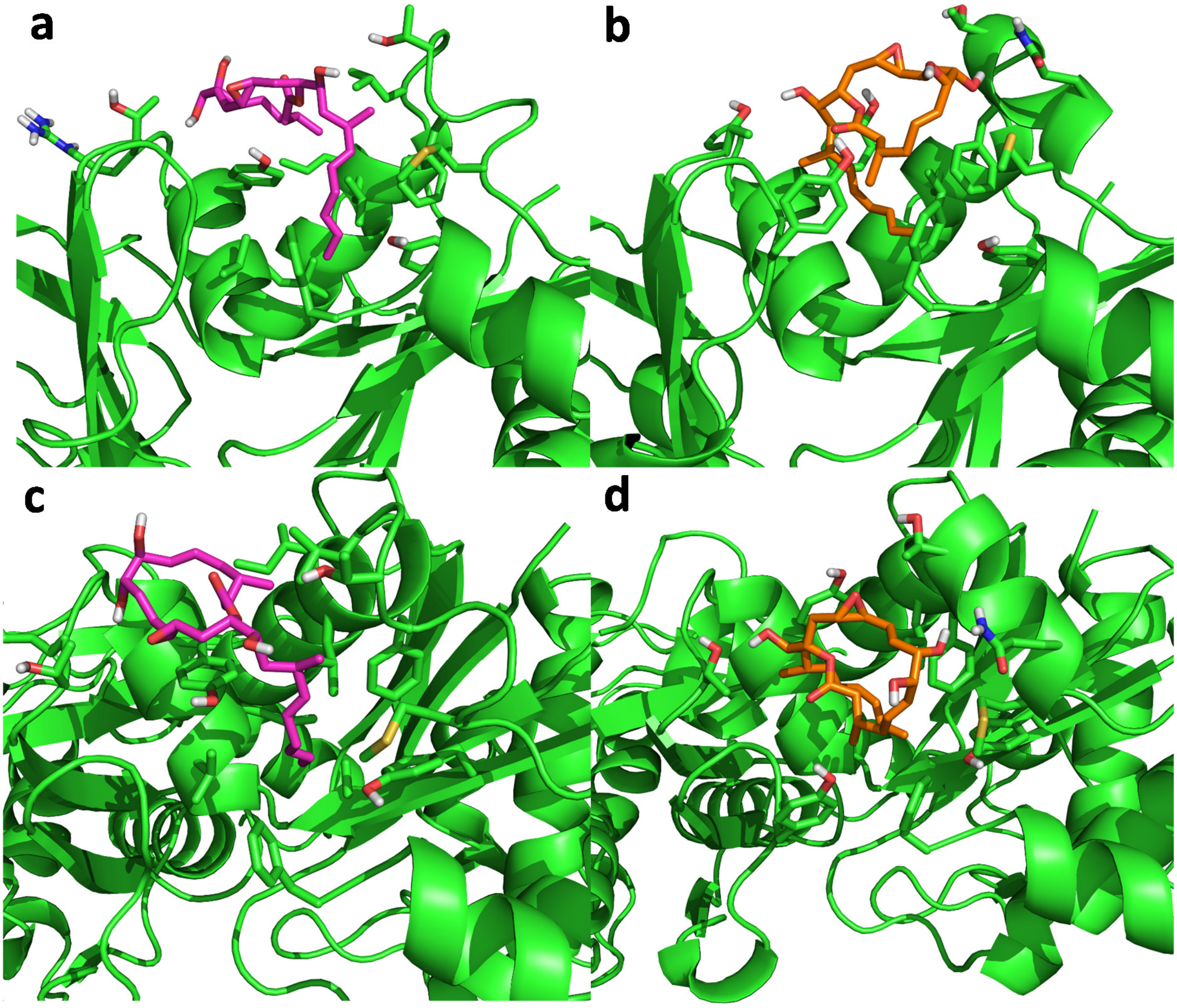


Figure 21: Cluster representatives of the two free energy basins sampled upon binding of iriomoteolide at the barbed end of monomeric G-actin. **a, c)** Cluster 1 was the largest cluster and the cluster representative shows a binding mode similar to natural products like reidispongiolide A. The hydrophobic tail is inserted into the barbed and the macrocycle is located along the hydrophobic patch. **b, d)** The representative of cluster 2 shows a similar arrangement of the hydrophobic tail while contrary to cluster 1 the macrolide resides on top of the tail section and not along the hydrophobic patch of the barbed end periphery. Images rendered in Pymol[19].

the chosen target cluster indicating robustness of the cFEP (Figure 19 a,b). A major difference between the original analyses and the supplemented one is that the two largest clusters are now in different basins. Interestingly, as with the 3x700 ns simulations the 2nd generation analysis shows that the largest cluster was sampled in fewer simulations (1x700 ns + 1x300 ns) but with higher total number of cluster members indicating higher stability of the system once the conformation is reached (Figure 21). Cluster 2 was sampled in more simulations (4x300 ns + 2x700 ns) but with overall less members (Figure 20, Supplementary table 5) which is characteristic for less stable binding modes. This was further

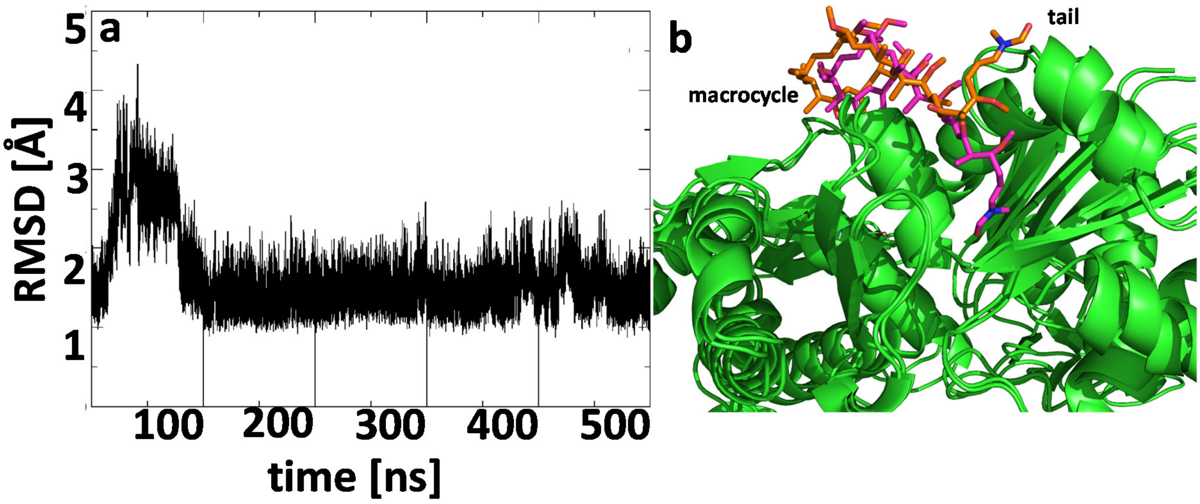


Figure 22: Overview of structural behavior of reidispongiolide A in five 100 ns simulations. **a)** RSMD of all heavy atoms of redA with the X-ray coordinates as reference. The first simulation shows an increased RMSD up to 4.3 Å while all other simulations show the RMSD fluctuating around 1.5 Å. **b)** Superposition of the reference crystal structure (magenta sticks) and the snapshot with the largest RMSD of 4.3 Å (orange sticks). This snapshot shows the dissociation of the tail section out of the barbed end while the marcocycle remained in its position. The overall shift resulted from the alignment being performed on all α-carbons and their respective movement compared to the reference snapshot. Figure generated in Pymol[19].

supplemented by the cluster progression displaying a less solid lines for cluster 2 compared to cluster 1. In conjunction with iriomoteolide not dissociating from the barbed end after the initial binding has happened, these results strongly indicated that iriomoteolide might be able to adopt several conformations in the barbed end of monomeric actin and that the mode of binding and its relative stability are strongly dependent on the conformation upon binding. The three dimensional arrangement of the cluster representatives with actin, revealed a binding mode similar to other macrolide natural products (Figure 21 a-d, 10a-e). Both clusters show an overall similar interaction with the protein whereas hydrophobic tail interacted with hydrophobic residues in the barbed end and the macrocycle located on the periphery. The most prominent difference between cluster 1 and cluster 2 is that the macro cycle in cluster 2 is located above the tail in a more condensed conformation (Figure 21 a-d). Cluster 1 represents a more stretched conformation with the macro cycle and hydrophobic tail providing a similar amount of interface with the barbed end. The closest residues (within 4.5 Å) of iriomoteolide with G-actin in cluster 1 were Phe375, Met355, Phe352, Thr351, Leu349, Leu346, Ile345, Gly168, Thr148, Arg147, Gly146, Tyr143, Leu140, A1a35 and Tyr 133. No hydrogen bond

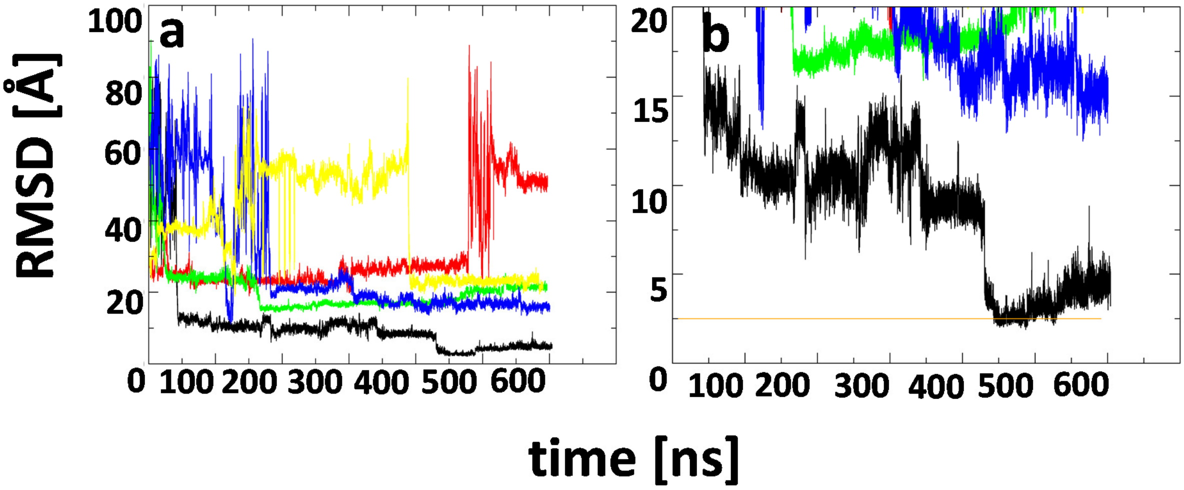


Figure 23: Time series for 5 x 600 ns simulations of redA with G-actin started from a dissociated state. **a)** Four of five simulations show RMSD values larger than 10 Å. **b)** Close up of the RMSD values with a base line set at 2.5 Å (orange line). Only a single simulation came close to sampling the crystal structure as observed in the simulations started from the crystal structure. All other simulations did not sample conformations with RMSD values below 10 Å.

geometries were recorded for any residue and iriomoteolide. Cluster 2 interacts with residues Phe375, Met355, Thr351, Gly168, Thr148, Tyr143, Val139 and Tyr133 displaying overall less contacts with the tail being the primary contribution to interaction with G-actin with no hydrogen bonds formed.

5.3.7 Simulations of bound reidispongiolide A support proposed mechanism of fibril disruption.

Five simulations carried out for bound reidispongiolide A (PDB-id: 2ASN) showed that over 100 ns a total of four simulations show a RMSD fluctuating between 1.5 Å and 2.5 Å over the entire simulations. Within this range redA is in contact with Tyr133, Val139, Ser141, Tyr143, Ala144, Ser145, Gly146, Arg147, Thr148, Gly168, Tyr169, Ile330, Pro332, Glu334 and Ser338, Ile341, Ile344, Leu345, Leu349, thr351, Met355 (Figure 11b,c). A single simulation showed an increase of RMSD up to 4.3 Å for all heavy atoms of redA (Figure 22a). Closer inspection of the reidispongiolide A conformation at the largest RMSD showed that the tail section is actually removed from the barbed end while the macro cycle stayed in place (Figure 22b). Upon rearrangement of the tail, interaction with Tyr133, Val139, Gly168, Tyr169, Glu334, Met355 is lost. The return of the RMSD below 3 Å was indicative for the tail being reinserted into the barbed end. Albeit this behavior being sampled only a single time in five simulations, it would support the proposed mechanism of the macro cycle providing the interface which interacts with G-actin first while the tail section is then inserted into the fibril severing it[48].

5.3.8 Binding simulations of redA do not reproduce crystal structure conformation on a 600 ns scale.

To assess the ability of molecular dynamics simulations being able to recreate the crystal structure conformation of redA, five simulations for 600 ns were started with random positions of redA in the active site. Compared to the binding simulations of iriomoteolide binding to the barbed end only a single simulation resulted in redA associating with the barbed end of the actin monomer in a conformation closely resembling the crystal conformation (Figure 23 a,b). For this simulation the RMSD gradually lowers until after 470 ns an RMSD of 3 Å is reached. Simulation 2 shows fast association of the redA with G-actin at a position outside of the barbed end with a RMSD of 22 Å. For 450 ns this position is maintained until dissociation at 500 ns and re-association with G-Actin 50 ns later shown by significant fluctuation until an RMSD of 50 Å was reached until the end of the simulation. Contrary, simulation 3 shows a redA being unbound for 30 ns after which an association event occurs and redA did not unbind from its site of interaction. Simulation 4 shows similar behavior with a dissociated phase being simulated for 230 ns after which it maintains a position different from the reference crystal structure conformation around 20 Å. Simulation 5 sampled redA for the longest period of time in an unbound state until at 450 ns it binds to the monomer with an RMSD similar to simulation 3.

The results show that the binding of redA to G-actin did not converge on a time scale of 600 ns. Compared to iriomoteolide binding to the barbed end was only observed in a single simulation for ~ 50 ns while for iriomoteolide three simulations sampled the inhibitor 1.75 µs associated with the barbed end (Figure 15). Similarly to iriomoteolide, redA mostly remains at the individual binding sites after the initial binding has occurred.

5.4 Discussion

In this study, the simulations of two naturally occurring macrolides iriomoteolide 3a and reidispongiolide A with G-actin were presented. Ten 100 ns binding simulations for an iriomoteolide to G-actin ratio of 10:1 showed high aggregation of iriomoteolide molecules in all simulations (Figure 13). Despite iriomoteolide providing three hydroxyl groups as well as an ester group and epoxide group, the propensity to aggregate in water resulted in at least three iriomoteolide molecules being closely associated (Figure 12a,13a-i). An interface for the aggregation of iriomoteolides with one another is most likely mediated by the hydrophobic tail of iriomoteolide as well as the aliphatic and methyl domains of the macro cycle. The aggregation of peptides into oligomers and lipids into micelles are processes well studied by molecular dynamics situation[94–96]. Given the accuracy of protein and lipid force fields to confidently model these such systems, it is unsurprising that iriomoteolide 3a would aggregate considering the large aliphatic tail section. However, iriomoteolide 3a provided three hydrogen bond donors and 6 hydrogen bond acceptors resulting in a total of 9 moieties for polar interaction with water. It is recommended to measure the *in vitro* aggregation of iriomoteolide 3a to reconcile the observation with the observed behavior *in silico* as it has been found out that aggregator molecules significantly influence polymerization of proteins[97].

The evaluation of contact frequencies showed a total of three hotspots of interaction surrounding the residues Phe352, Met283, and Met355 out of which Phe352 represented a central residue of the barbed end (Figure 15) close to Gly168. Binding of iriomoteolide to the barbed end of monomeric actin would provide the most plausible region of polymerization suppression as a multitude of other natural products interacts with the barbed end of actin[27]. However, considering the number of aggregates, it was naturally assumed that the observed contact frequencies are influenced by aggregates artificially, increasing residence times of an iriomoteolide interacting with G-actin.

To avoid artifacts in binding behavior, a series of simulations was carried out with an iriomoteolide to actin ratio of 1:1. Five simulations of 700 ns each showed three different sites of residence (Met283, Met355, Phe352) with the barbed end (and representative Phe352/Gly168) being the prime interaction site which was sampled three times. Binding of iriomoteolide happened within 200 ns in all three cases and resulted in a cumulative residence time of 1.75 µs at the barbed end. Two simulations resulted in sampling of the secondary interaction sites around Met283 and Met355. These sites are identical to the high concentration iriomoteolide simulations. The main difference between the results is the magnitude of relative contact frequencies. While for the 10x100 ns simulations, involving 10 iriomoteolide molecules, Met355 was the site of highest contact frequency Phe352 was the residue of highest contact frequency for the 5x700 ns simulations (Supplementary Table 1,2). It is interesting to point out that in both simulation schemes, the three sites of highest contact frequency are around the same residues giving some level of confidence to the observed interactions. Out of all three interaction sites the barbed end with Gly168 (close to Phe352) would represent the viable site for inhibiting G-actin polymerization. Several crystal structures for natural product macrolides have already shown that these molecules bind at the barbed end while having in vitro properties that lead to the sequestering of monomers, fibril disruption and polymerization inhibition[47–51]. To asses if the system was equilibrated a series of 25 simulations was run for 100 ns and 5 different positions of iriomoteolide in the box. These simulation show a very similar behavior of iriomoteolide associating with G-actin. Identical to the previous two simulation schemes, it was observed that three primary interaction hotspots are around residues Gly168, Met283, and Met355. The combined total of 5.6 µs in which interaction with the barbed end was sampled the most frequent and the three primary interaction sites being identical, indicated an equilibrated system which shows that iriomoteolide binds preferably at the barbed end.

Finally, it is worthwhile highlighting that upon binding towards the barbed end iriomoteolide did not dissociate for the remainder of the respective simulation which lead to further confidence that the barbed end is the prime site of actin polymerization inhibition by iriomoteolide.

To obtain further structural insight into the specific interaction of iriomoteolide with the barbed end all simulations where binding of iriomoteolide to the barbed end was observed were extended to simulations exceeding 500 ns each to a combined total of 1.75 µs. Iriomoteolide was then clustered and a cut-based free energy profile was generated to assess conformations upon binding. The cut-based free energy profile for the combined simulations revealed two distinct basins showing two distinct modes of binding. Both are similar in their relative arrangement with the hydrophobic tail being inserted into the barbed end interacting with its hydrophobic patch while the macro cycle is located either on top of the tail in what is considered a condensed conformation. An extended conformation is characterized by the macrocycle being located on the periphery of the barbed end similar to macrolides like reidispongiolide A and trisoxasoles[47].

Out of the two conformations, the extended conformation was more stable but occurring in less simulations while the condensed conformation occurred in more simulations but was less often sampled (Figure 20). From this data, at least two possible modes of interaction could be assumed as no transition from one binding mode to the other was observed in the simulation. A comparison the binding modes with X-ray crystallography data from reidispongiolide A bound to G-actin showed that the extended conformation is the one closer to redA and other natural products as the number of residues which are located within 4 Å of irio3a and redA is high for the extended conformation. Finally, the author suggests that in order to fully validate the findings presented here, NMR spectroscopy or X-ray crystallography should be carried out.

The simulations of bound reidispongiolide A suggest that the binding of the macrocycle is stronger than the hydrophobic tail. The single observed tail dissociation would support the mode of interaction as proposed by Allingham et al. 2005 whereas the macrocycle interacts first with the fiber and then inserts the tail into the barbed end, severing the filament[48]. Binding simulations of reidispongiolide A were not able to sufficiently reproduce the crystal structure indicating the need for extended simulations. None of the 5 simulations of unbound redA was able to generate binding behavior similar to the simulations of bound reidispongiolide A.

5.5 References

1. Korn ED. Actin polymerization and its regulation by proteins from nonmuscle cells. *Physiol Rev*. 1982;62(2):672-737.

2. Fletcher DA, Mullins RD. Cell mechanics and the cytoskeleton. *Nature*. 2010;463(7280):485-492. doi:10.1038/nature08908.

3. Wickstead B, Gull K. The evolution of the cytoskeleton. *J Cell Biol*. 2011;194(4):513-525. doi:10.1083/jcb.201102065.

4. Chrétien D, Fuller SD, Karsenti E. Structure of growing microtubule ends: two-dimensional sheets close into tubes at variable rates. *J Cell Biol*. 1995;129(5):1311-1328. doi:10.1083/jcb.129.5.1311.

5. Nogales E. Structural Insights into Microtubule Function. *Annu Rev Biochem*. 2000;69(1):277-302. doi:10.1146/annurev.biochem.69.1.277.

6. Desai A, Mitchison TJ. Microtubule Polymerization Dynamics. *Annu Rev Cell Dev Biol*. 1997;13(1):83-117. doi:10.1146/annurev.cellbio.13.1.83.

7. Burbank KS, Mitchison TJ. Microtubule dynamic instability. *Curr Biol*. 2006;16(14):R516-R517. doi:10.1016/j.cub.2006.06.044.

8. Rieder CL, Salmon ED. The vertebrate cell kinetochore and its roles during mitosis. *Trends Cell Biol*. 1998;8(8):310-318. doi:10.1016/S0962-8924(98)01299-9.

9. Mitchison T, Kirschner M. Dynamic instability of microtubule growth. *Nature*. 1984;312(5991):237-242. doi:10.1038/312237a0.

10. David-Pfeuty T, Erickson HP, Pantaloni D. Guanosinetriphosphatase activity of tubulin associated with microtubule assembly. *Proc Natl Acad Sci U S A*. 1977;74(12):5372-5376.

11. Drechsel DN, Kirschner MW. The minimum GTP cap required to stabilize microtubules. *Curr Biol*. 1994;4(12):1053-1061. doi:10.1016/S0960-9822(00)00243-8.

12. Caplow M, Shanks J. Evidence that a single monolayer tubulin-GTP cap is both necessary and sufficient to stabilize microtubules. *Mol Biol Cell*. 1996;7(4):663-675.

13. Vale RD. The Molecular Motor Toolbox for Intracellular Transport. *Cell*. 2003;112(4):467-480. doi:10.1016/S0092-8674(03)00111-9.

14. Herrmann H, Aebi U. INTERMEDIATE FILAMENTS: Molecular Structure, Assembly Mechanism, and Integration Into Functionally Distinct Intracellular Scaffolds. *Annu Rev Biochem*. 2004;73(1):749-789. doi:10.1146/annurev.biochem.73.011303.073823.

15. Goldman RD, Khuon S, Chou YH, Opal P, Steinert PM. The function of intermediate filaments in cell shape and cytoskeletal integrity. *J Cell Biol*. 1996;134(4):971-983.

16. Fuchs E, Weber K. Intermediate Filaments: Structure, Dynamics, Function and Disease. *Annu Rev Biochem*. 1994;63(1):345-382. doi:10.1146/annurev.bi.63.070194.002021.

17. Herrmann H, Bär H, Kreplak L, Strelkov SV, Aebi U. Intermediate filaments: from cell architecture to nanomechanics. *Nat Rev Mol Cell Biol*. 2007;8(7):562-573. doi:10.1038/nrm2197.

18. Ayaz P, Ye X, Huddleston P, Brautigam CA, Rice LM. A TOG:αβ-tubulin Complex Structure Reveals Conformation-Based Mechanisms for a Microtubule Polymerase. *Science*. 2012;337(6096):857-860. doi:10.1126/science.1221698.

19. The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

20. Crick FHC. Is α-Keratin a Coiled Coil? *Nature*. 1952;170(4334):882-883. doi:10.1038/170882b0.

21. Omary MB, Coulombe PA, McLean WHI. Intermediate Filament Proteins and Their Associated Diseases. *N Engl J Med*. 2004;351(20):2087-2100. doi:10.1056/NEJMra040319.

22. Nicolet S, Herrmann H, Aebi U, Strelkov SV. Atomic structure of vimentin coil 2. *J Struct Biol*. 2010;170(2):369-376. doi:10.1016/j.jsb.2010.02.012.

23. Pollard TD, Cooper JA. Actin, a Central Player in Cell Shape and Movement. *Science*. 2009;326(5957):1208-1212. doi:10.1126/science.1175862.

24. Pollard TD, Blanchoin L, Mullins RD. Molecular Mechanisms Controlling Actin Filament Dynamics in Nonmuscle Cells. *Annu Rev Biophys Biomol Struct*. 2000;29(1):545-576. doi:10.1146/annurev.biophys.29.1.545.

25. Pollard TD, Cooper JA. Actin and Actin-Binding Proteins. A Critical Evaluation of Mechanisms and Functions. *Annu Rev Biochem*. 1986;55(1):987-1035. doi:10.1146/annurev.bi.55.070186.005011.

26. Oda T, Iwasa M, Aihara T, Maéda Y, Narita A. The nature of the globular- to fibrous-actin transition. *Nature*. 2009;457(7228):441-445. doi:10.1038/nature07685.

27. Allingham JS, Klenchin VA, Rayment I. Actin-targeting natural products: structures, properties and mechanisms of action. *Cell Mol Life Sci CMLS*. 2006;63(18):2119-2134. doi:10.1007/s00018-006-6157-9.

28. Perrin BJ, Ervasti JM. The actin gene family: Function follows isoform. *Cytoskeleton*. 2010;67(10):630-634. doi:10.1002/cm.20475.

29. Herman IM. Actin isoforms. *Curr Opin Cell Biol*. 1993;5(1):48-55.

30. Kudryashov DS, Reisler E. ATP and ADP actin states. *Biopolymers*. 2013;99(4):245-256. doi:10.1002/bip.22155.

31. Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC. Atomic structure of the actin: DNase I complex. *Nature*. 1990;347(6288):37-44. doi:10.1038/347037a0.

32. Splettstoesser T, Noé F, Oda T, Smith JC. Nucleotide-dependence of G-actin conformation from multiple molecular dynamics simulations and observation of a putatively polymerization-competent superclosed state. *Proteins Struct Funct Bioinforma*. 2009;76(2):353-364. doi:10.1002/prot.22350.

33. Holmes KC. Structural biology: Actin in a twist. *Nature*. 2009;457(7228):389-390. doi:10.1038/457389a.

34. Korn ED, Carlier MF, Pantaloni D. Actin polymerization and ATP hydrolysis. *Science*. 1987;238(4827):638-644. doi:10.1126/science.3672117.

35. Hanson J, Lowy J. The structure of F-actin and of actin filaments isolated from muscle. *J Mol Biol*. 1963;6(1):46-IN5. doi:10.1016/S0022-2836(63)80081-9.

36. Holmes KC, Popp D, Gebhard W, Kabsch W. Atomic model of the actin filament. *Nature*. 1990;347(6288):44-49. doi:10.1038/347044a0.

37. Holmes KC. 50 years of fiber diffraction. *J Struct Biol*. 2010;170(2):184-191. doi:10.1016/j.jsb.2010.01.004.

38. Oda T, Maéda Y. Multiple Conformations of F-actin. *Structure*. 2010;18(7):761-767. doi:10.1016/j.str.2010.05.009.

39. Krishnan K, Moens PDJ. Structure and functions of profilins. *Biophys Rev*. 2009;1(2):71-81. doi:10.1007/s12551-009-0010-y.

40. Goode BL, Eck MJ. Mechanism and Function of Formins in the Control of Actin Assembly. *Annu Rev Biochem*. 2007;76(1):593-627. doi:10.1146/annurev.biochem.75.103004.142647.

41. McLaughlin PJ, Gooch JT, Mannherz H-G, Weeds AG. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. *Nature*. 1993;364(6439):685-692. doi:10.1038/364685a0.

42. Oda T, Namba K, Maéda Y. Position and Orientation of Phalloidin in F-Actin Determined by X-Ray Fiber Diffraction Analysis. *Biophys J*. 2005;88(4):2727-2736. doi:10.1529/biophysj.104.047753.

43. Oda T, Crane ZD, Dicus CW, Sufi BA, Bates RB. Dolastatin 11 Connects Two Long-pitch Strands in F-actin to Stabilize Microfilaments. *J Mol Biol*. 2003;328(2):319-324. doi:10.1016/S0022-2836(03)00306-1.

44. Usui T, Kazami S, Dohmae N, et al. Amphidinolide H, a Potent Cytotoxic Macrolide, Covalently Binds on Actin Subdomain 4 and Stabilizes Actin Filament. *Chem Biol*. 2004;11(9):1269-1277. doi:10.1016/j.chembiol.2004.07.014.

45. Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology*. Vol 4th ed. W. H. Freeman

46. Sept D, McCammon JA. Thermodynamics and Kinetics of Actin Filament Nucleation. *Biophys J*. 2001;81(2):667-674. doi:10.1016/S0006-3495(01)75731-1.

47. Klenchin VA, Allingham JS, King R, Tanaka J, Marriott G, Rayment I. Trisoxazole macrolide toxins mimic the binding of actin-capping proteins to actin. *Nat Struct Mol Biol*. 2003;10(12):1058-1063. doi:10.1038/nsb1006.

48. Allingham JS, Zampella A, D'Auria MV, Rayment I. Structures of microfilament destabilizing toxins bound to actin provide insight into toxin design and activity. *Proc Natl Acad Sci U S A*. 2005;102(41):14527-14532. doi:10.1073/pnas.0502089102.

49. Klenchin VA, King R, Tanaka J, Marriott G, Rayment I. Structural Basis of Swinholide A Binding to Actin. *Chem Biol*. 2005;12(3):287-291. doi:10.1016/j.chembiol.2005.02.011.

50. Hirata K, Muraoka S, Suenaga K, et al. Structure Basis for Antitumor Effect of Aplyronine A. *J Mol Biol*. 2006;356(4):945-954. doi:10.1016/j.jmb.2005.12.031.

51. Rizvi SA, Tereshko V, Kossiakoff AA, Kozmin SA. Structure of Bistramide A–Actin Complex at a 1.35 Å Resolution. *J Am Chem Soc*. 2006;128(12):3882-3883. doi:10.1021/ja058319c.

52. Bamburg JR, Bernstein BW. Roles of ADF/cofilin in actin polymerization and beyond. *F1000 Biol Rep*. 2010;2. doi:10.3410/B2-62.

53. Elam WA, Kang H, De La Cruz EM. Biophysics of actin filament severing by cofilin. *FEBS Lett.* 2013;587(8):1215-1219. doi:10.1016/j.febslet.2013.01.062.

54. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 1995;8(2):127-134.

55. Spector I, Braet F, Shochet NR, Bubb MR. New anti-actin drugs in the study of the organization and function of the actin cytoskeleton. *Microsc Res Tech.* 1999;47(1):18-37. doi:10.1002/(SICI)1097-0029(19991001)47:1<18::AID-JEMT3>3.0.CO;2-E.

56. Coué M, Brenner SL, Spector I, Korn ED. Inhibition of actin polymerization by latrunculin A. *FEBS Lett.* 1987;213(2):316-318. doi:10.1016/0014-5793(87)81513-2.

57. Morton WM, Ayscough KR, McLaughlin PJ. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. *Nat Cell Biol.* 2000;2(6):376-378. doi:10.1038/35014075.

58. Yarmola EG, Somasundaram T, Boring TA, Spector I, Bubb MR. Actin-latrunculin A structure and function:differential modulation of actin-binding protein function by latrunculin A. *J Biol Chem.* June 2000. doi:10.1074/jbc.M004253200.

59. Rennebaum S, Caflisch A. Inhibition of interdomain motion in g-actin by the natural product latrunculin: A molecular dynamics study. *Proteins Struct Funct Bioinforma.* 2012;80(8):1998-2008. doi:10.1002/prot.24088.

60. Tsuda M, Oguchi K, Iwamoto R, et al. Iriomoteolide-1a, a Potent Cytotoxic 20-Membered Macrolide from a Benthic Dinoflagellate Amphidinium Species. *J Org Chem.* 2007;72(12):4469-4474. doi:10.1021/jo070414b.

61. Tsuda M, Oguchi K, Iwamoto R, et al. Iriomoteolides-1b and -1c, 20-Membered Macrolides from a Marine Dinoflagellate Amphidinium Species. *J Nat Prod.* 2007;70(10):1661-1663. doi:10.1021/np0702537.

62. Oguchi K, Tsuda M, Iwamoto R, et al. Iriomoteolide-3a, a Cytotoxic 15-Membered Macrolide from a Marine Dinoflagellate Amphidinium Species. *J Org Chem.* 2008;73(4):1567-1570. doi:10.1021/jo702440s.

63. Cribiú R, Jäger C, Nevado C. Syntheses and Biological Evaluation of Iriomoteolide 3a and Analogues. *Angew Chem Int Ed.* 2009;48(46):8780-8783. doi:10.1002/anie.200903379.

64. Unzue A, Nevado C. *Master's Thesis: Iriomoteolides as Novel Chemical Tools for the Study of Actin Dynamics.* University of Zurich; 2012.

65. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem.* 1983;4(2):187-217. doi:10.1002/jcc.540040211.

66. MacKerell AD, Brooks B, Brooks CL, et al. CHARMM: The Energy Function and Its Parameterization. In: *Encyclopedia of Computational Chemistry.* Vol John Wiley & Sons, Ltd; 2002. http://onlinelibrary.wiley.com/doi/10.1002/0470845015.cfa007/abstract. Accessed March 28, 2015.

67. Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: The biomolecular simulation program. *J Comput Chem.* 2009;30(10):1545-1614. doi:10.1002/jcc.21287.

68. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. 1995;91(1–3):43-56. doi:10.1016/0010-4655(95)00042-E.

69. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-1718. doi:10.1002/jcc.20291.

70. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput*. 2010;6(2):459-466. doi:10.1021/ct900549r.

71. Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem*. 2004;25(11):1400-1415. doi:10.1002/jcc.20065.

72. Clark M, Cramer RD, Van Opdenbosch N. Validation of the general purpose tripos 5.2 force field. *J Comput Chem*. 1989;10(8):982-1012. doi:10.1002/jcc.540100804.

73. Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010;31(4):671-690. doi:10.1002/jcc.21367.

74. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J Chem Inf Model*. 2012;52(12):3144-3154. doi:10.1021/ci300363c.

75. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J Chem Inf Model*. 2012;52(12):3155-3168. doi:10.1021/ci3003649.

76. Yu W, He X, Vanommeslaeghe K, MacKerell AD. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem*. 2012;33(31):2451-2468. doi:10.1002/jcc.23067.

77. GROMACS Manual 4.5.6 Chapter 5 (Topologies).

78. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926. doi:10.1063/1.445869.

79. Toukan K, Rahman A. Molecular-dynamics study of atomic motions in water. *Phys Rev B*. 1985;31(5):2643-2648. doi:10.1103/PhysRevB.31.2643.

80. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. *J Chem Phys*. 1996;105(5):1902-1921. doi:10.1063/1.472061.

81. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):014101. doi:10.1063/1.2408420.

82. Bussi G, Parrinello M. Stochastic thermostats: comparison of local and global schemes. *Comput Phys Commun*. 2008;179(1–3):26-29. doi:10.1016/j.cpc.2008.01.006.

83. Bussi G, Zykova-Timan T, Parrinello M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J Chem Phys*. 2009;130(7):074101. doi:10.1063/1.3073889.

84. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089-10092. doi:10.1063/1.464397.

85. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19):8577-8593. doi:10.1063/1.470117.

86. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*. 1997;18(12):1463-1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

87. Krivov SV, Karplus M. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J Phys Chem B*. 2006;110(25):12689-12698. doi:10.1021/jp060039b.

88. Krivov SV, Karplus M. Diffusive reaction dynamics on invariant free energy profiles. *Proc Natl Acad Sci*. 2008;105(37):13841-13846. doi:10.1073/pnas.0800228105.

89. Vitalis A, Caflisch A. Efficient Construction of Mesostate Networks from Molecular Dynamics Trajectories. *J Chem Theory Comput*. 2012;8(3):1108-1120. doi:10.1021/ct200801b.

90. Vitalis A, Pappu RV. Chapter 3 Methods for Monte Carlo Simulations of Biomacromolecules. In: Wheeler RA, ed. *Annual Reports in Computational Chemistry*. Vol 5. Elsevier; 2009:49-76. http://www.sciencedirect.com/science/article/pii/S1574140009005039. Accessed March 28, 2015.

91. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*. 2007;23(19):2625-2627. doi:10.1093/bioinformatics/btm378.

92. Seeber M, Felline A, Raimondi F, et al. Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem*. 2011;32(6):1183-1194. doi:10.1002/jcc.21688.

93. Krivov SV, Muff S, Caflisch A, Karplus M. One-dimensional barrier-preserving free-energy projections of a beta-sheet miniprotein: new insights into the folding process. *J Phys Chem B*. 2008;112(29):8701-8714. doi:10.1021/jp711864r.

94. Paci E, Gsponer J, Salvatella X, Vendruscolo M. Molecular Dynamics Studies of the Process of Amyloid Aggregation of Peptide Fragments of Transthyretin. *J Mol Biol*. 2004;340(3):555-569. doi:10.1016/j.jmb.2004.05.009.

95. Marrink SJ, Tieleman DP, Mark AE. Molecular Dynamics Simulation of the Kinetics of Spontaneous Micelle Formation. *J Phys Chem B*. 2000;104(51):12165-12173. doi:10.1021/jp001898h.

96. Böckmann RA, Caflisch A. Spontaneous formation of detergent micelles around the outer membrane protein OmpX. *Biophys J*. 2005;88(5):3191-3204. doi:10.1529/biophysj.105.060426.

97. Rishton GM. Aggregator compounds confound amyloid fibrillization assay. *Nat Chem Biol*. 2008;4(3):159-160. doi:10.1038/nchembio0308-159.

5.6 Supplementary material

residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i
6THR	40	58ALA	2	98PRO	3	147ARG	21	196ARG	17	234SER	5	272ALA	12	317ILE	10	354GLN	43
7ALA	33	59GLN	12	99GLU	5	148THR	48	197GLY	17	235SER	7	273GLY	4	318THR	23	355MET	60
19ALA	1	60SER	14	100GLU	10	149THR	3	198TYR	4	236LEU	10	275HSD	10	319ALA	19	356TRP	29
20GLY	4	61LYS	9	101HSD	19	161HSD	1	199SER	21	237GLU	3	276GLU	30	320LEU	36	357ILE	1
21PHE	17	62ARG	24	102PRO	30	165ILE	1	200PHE	10	238LYS	6	277THR	19	321ALA	35	358THR	17
22ALA	16	63GLY	19	109PRO	19	166TYR	7	201VAL	36	239SER	6	278THR	5	322PRO	46	359LYS	5
23GLY	16	64ILE	10	110LEU	35	167GLU	50	202THR	34	240TYR	3	279TYR	41	323SER	18	360GLN	6
24ASP	17	65LEU	2	111ASN	23	168GLY	73	203THR	35	241GLU	2	280ASN	63	324THR	30	361GLU	3
25ASP	6	66THR	11	112PRO	29	169TYR	58	204ALA	21	242LEU	1	281SER	14	325MET	39	362TYR	1
26ALA	5	67LEU	8	113LYS	16	170ALA	17	205GLU	9	243PRO	7	282ILE	10	326LYS	22	363ASP	1
27PRO	5	68LYS	5	114ALA	6	171LEU	5	206ARG	14	244ASP	2	283MET	81	327ILE	13	364GLU	5
28ARG	15	69TYR	5	115ASN	5	172PRO	33	207GLU	8	245GLY	2	284LYS	44	328LYS	16	365ALA	9
29ALA	5	72GLU	4	116ARG	1	173HSD	37	208ILE	6	246GLN	3	285CYS	6	329ILE	9	366GLY	4
30VAL	3	73HIM	6	117GLU	1	174ALA	19	210ARG	1	247VAL	2	286ASP	7	330ILE	3	367PRO	2
37ARG	4	74GLY	2	118LYS	2	175ILE	32	211ASP	2	248ILE	1	287ILE	30	331ALA	3	368SER	10
38PRO	8	75ILE	6	121GLN	1	176MET	28	214GLU	1	249THR	2	288ASP	9	332PRO	6	369ILE	9
39ARG	9	76ILE	3	125GLU	1	177ARG	20	215LYS	6	251GLY	12	289ILE	1	333PRO	4	370VAL	4
40HSD	9	77THR	4	126THR	3	178LEU	4	216LEU	6	252ASN	48	290ARG	30	334GLU	2	371HSD	17
41GLN	13	78ASN	1	127PHE	3	179ASP	8	217CYS	3	253GLU	14	291LYS	16	337TYR	1	372ARG	32
42GLY	15	79TRP	1	128ASN	1	180LEU	5	218TYR	5	254ARG	4	292ASP	5	338SER	1	373LYS	34
43VAL	16	80ASP	1	129VAL	2	181ALA	1	221LEU	6	255PHE	5	294TYR	11	341ILE	7	374CYS	24
44MET	13	83GLU	3	130PRO	24	183ARG	3	222ASP	5	256ARG	52	295ALA	4	342GLY	16	375PHE	49
45VAL	12	84LYS	4	131ALA	29	184ASP	8	223PHE	56	259GLU	5	296ASN	3	343GLY	3		
46GLY	15	87HSD	23	133TYR	5	186THR	4	224GLU	75	260THR	1	305MET	1	344SER	7		
47MET	15	88HSD	8	135ALA	3	187ASP	18	225ASN	26	263GLN	14	306TYR	1	345ILE	35		
48GLY	16	90PHE	2	136ILE	4	188TYR	11	226GLU	2	264PRO	4	307PRO	5	346LEU	45		
49GLN	14	91TYR	40	139VAL	20	189LEU	0	227MET	100	265SER	30	308GLY	1	347ALA	19		
50LYS	10	92ASN	30	140LEU	11	190MET	16	228ALA	80	266PHE	48	310ALA	3	348SER	20		
51ASP	13	93GLU	4	142LEU	17	191LYS	32	229THR	23	267ILE	26	311ASP	15	349LEU	51		
52SER	6	94LEU	8	143TYR	55	192ILE	11	230ALA	43	268GLY	27	312ARG	1	350SER	21		
53TYR	3	95ARG	40	144ALA	21	193LEU	1	231ALA	77	269MET	20	314GLN	20	351THR	63		
56ASP	3	96VAL	31	145SER	11	194THR	31	232SER	37	270GLU	11	315LYS	22	352PHE	64		
57GLU	1	97ALA	27	146GLY	38	195GLU	30	233SER	8	271SER	9	316GLU	6	353GLN	34		

Table 1: Relative contact frequencies (C_i) for all residues of the 10x100 ns scheme with a irio3a to actin ratio of 10 : 1. All residues having a C_i of 0 % were omitted.

residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i
6THR	23	94LEU	1	197GLY	1	283MET	2	354GLN	39
7ALA	23	95ARG	1	198TYR	1	284LYS	1	355MET	72
16LEU	1	96VAL	1	199SER	1	287ILE	1	356TRP	28
18LYS	1	100GLU	12	201VAL	2	290ARG	2	357ILE	1
20GLY	1	101HIS	18	202THR	1	291LYS	1	358THR	24
21PHE	1	102PRO	28	203THR	1	294TYR	1	359LYS	3
22ALA	1	103THR	1	206ARG	1	311ASP	1	360GLN	3
23GLY	1	129VAL	4	221LEU	1	314GLN	1	361GLU	2
24ASP	1	130PRO	25	222ASP	1	315LYS	2	364GLU	1
26ALA	1	131ALA	28	223PHE	24	318THR	1	372ARG	2
27PRO	2	133TYR	13	224GLU	24	319ALA	1	373LYS	19
28ARG	3	135ALA	1	225ASN	2	320LEU	1	374CYS	12
29ALA	2	136ILE	1	226GLU	1	321ALA	1	375PHE	53
30VAL	2	139VAL	61	227MET	27	322PRO	2		
39ARG	2	140LEU	29	228ALA	17	323SER	1		
40HIS	3	142LEU	27	229THR	3	324THR	1		
41GLN	1	143TYR	76	230ALA	13	325MET	2		
42GLY	2	144ALA	13	231ALA	15	326LYS	1		
43VAL	3	146GLY	27	232SER	3	327ILE	1		
44MET	2	147ARG	3	246GLN	1	336LYS	1		
45VAL	3	148THR	76	247VAL	1	337TYR	1		
46GLY	3	149THR	2	252ASN	16	342GLY	23		
47MET	3	165ILE	1	255PHE	5	343GLY	2		
48GLY	2	167GLU	58	256ARG	22	345ILE	28		
49GLN	1	168GLY	98	259GLU	7	346LEU	90		
61LYS	1	169TYR	63	263GLN	7	347ALA	15		
63GLY	2	170ALA	1	265SER	7	348SER	4		
64ILE	3	187ASP	1	266PHE	16	349LEU	60		
65LEU	1	190MET	1	276GLU	1	350SER	11		
66THR	1	191LYS	1	279TYR	2	351THR	92		
67LEU	1	194THR	1	280ASN	2	352PHE	100		
68LYS	1	195GLU	1	282ILE	1	353GLN	29		

Table 2: Relative contact frequencies (C_i) for residues of the 5x700 ns scheme with a irio3a to actin ratio of 1 : 1. All residues having a C_i of 0 % were omitted.

residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i	residue	C_i
36GLY	1	115ASN	5	197GLY	19	250ILE	1	315LYS	29	367PRO	1
37ARG	3	117GLU	1	198TYR	11	251GLY	3	316GLU	2	368SER	6
38PRO	7	118LYS	1	199SER	17	252ASN	21	318THR	26	369ILE	7
39ARG	19	124PHE	3	200PHE	5	253GLU	3	319ALA	12	371HSD	1
40HSD	18	125GLU	2	201VAL	16	255PHE	3	320LEU	28	372ARG	9
41GLN	5	128ASN	5	202THR	10	256ARG	24	321ALA	26	373LYS	35
42GLY	7	133TYR	32	203THR	4	259GLU	2	322PRO	29	374CYS	22
43VAL	14	135ALA	6	204ALA	3	263GLN	5	323SER	8	375PHE	74
44MET	7	136ILE	10	205GLU	3	265SER	7	324THR	8		
45VAL	22	139VAL	55	206ARG	2	266PHE	11	325MET	17		
46GLY	27	140LEU	23	221LEU	17	270GLU	1	326LYS	11		
47MET	24	142LEU	5	222ASP	9	271SER	3	327ILE	11		
48GLY	15	143TYR	53	223PHE	28	272ALA	6	328LYS	3		
49GLN	8	148THR	56	224GLU	39	276GLU	10	329ILE	1		
60SER	6	149THR	3	225ASN	14	277THR	7	345ILE	1		
61LYS	17	165ILE	7	226GLU	1	279TYR	35	346LEU	80		
63GLY	19	166TYR	2	227MET	47	280ASN	41	349LEU	35		
64ILE	29	167GLU	62	228ALA	39	281SER	7	350SER	7		
65LEU	11	168GLY	98	229THR	9	282ILE	13	351THR	83		
66THR	4	169TYR	77	230ALA	20	283MET	43	352PHE	100		
67LEU	2	170ALA	8	231ALA	28	284LYS	17	353GLN	1		
68LYS	2	173HSD	1	232SER	13	286ASP	1	354GLN	29		
72GLU	2	174ALA	4	233SER	2	287ILE	7	355MET	90		
73HIM	2	175ILE	3	234SER	1	288ASP	2	356TRP	7		
75ILE	3	176MET	9	235SER	1	290ARG	17	358THR	3		
77THR	4	177ARG	4	237GLU	1	291LYS	2	359LYS	7		
78ASN	1	178LEU	1	242LEU	4	292ASP	1	360GLN	10		
79TRP	1	187ASP	4	244ASP	3	294TYR	6	361GLU	9		
110LEU	2	190MET	7	245GLY	6	308GLY	1	362TYR	4		
111ASN	2	191LYS	10	246GLN	11	310ALA	2	363ASP	7		
112PRO	6	194THR	15	247VAL	10	311ASP	21	364GLU	10		
113LYS	1	195GLU	8	248ILE	8	312ARG	1	365ALA	9		
114ALA	4	196ARG	6	249THR	4	314GLN	24	366GLY	4		

Table 3: Relative contact frequencies (C_i) for residues of the 25x100 ns scheme with a irio3a to actin ratio of 1 : 1. All residues having a C_i of 0 % were omitted.

#	No.	Center	Diameter	Radius
1	11634	123700	1.6681	1.1795
2	8414	103019	1.7713	1.2524
3	7444	39318	1.6487	1.1658
4	5198	112710	1.7623	1.246
5	5046	40788	1.5572	1.101
6	4562	28274	1.9498	1.3786
7	4131	136242	1.5732	1.1123
8	3651	206680	1.6021	1.1327
9	3040	177933	1.6171	1.1433
10	2715	96779	1.7415	1.2312

Table 4: Tree-based clustering results for the 3x700 ns simulations.

#	No.	Center	Diameter	Radius
1	11921	102759	1.5711	1.1109
2	9158	39318	1.6321	1.154
3	7242	310176	1.7423	1.2319
4	6651	350726	1.5256	1.0787
5	5795	136949	1.5803	1.1173
6	5762	244005	1.8068	1.2775
7	5089	309459	1.7628	1.2464
8	4967	264312	1.6562	1.171
9	4965	177933	1.5657	1.107
10	4461	124049	1.7342	1.2262

Table 5: Tree-based clustering results for the combined 3x700 ns + 5 x 300 ns simulations

```
title       = Protein-ligand
integrator  = md
nsteps      = 50000000
dt          = 0.002
; Output control
nstxout     = 0000
nstvout     = 0000
nstenergy   = 0000
nstlog      = 5000
nstxtcout   = 5000
energygrps  = Protein ATP IRIO
; Bond parameters
continuation    = no
constraint_algorithm = lincs
constraints     = hbonds
lincs_iter      = 1
lincs_order     = 4
; Neighborsearching
ns_type     = grid
nstlist     = 5
rlist       = 1.0
;Van der Walls
vdwtype     = Cut-Off
rvdw        = 1.0
; Electrostatics
coulombtype     = PME
rcoulomb = 1.0
pme_order       = 4
fourierspacing  = 0.12
; Temperature coupling is on
tcoupl      = V-rescale
tc-grps     = Protein_ATP_IRIO Water_and_ions
tau_t       = 0.1   0.1
ref_t       = 310   310
; Pressure coupling is on
pcoupl      = Parrinello-Rahman
pcoupltype  = isotropic
tau_p       = 2.0
ref_p       = 1.01325
compressibility = 4.5e-5
```

```
; Periodic boundary conditions
pbc         = xyz
; Dispersion correction
DispCorr    = EnerPres
; Velocity generation
gen_vel     = yes
gen_seed    = -1
gen_temp    = 310
```

Example .mdp file for the production simulations as presented in this study.

Chapter 6

Conclusions & outlook

The thesis presents four projects. While the study of cavitands and actin/iriomoteolide 3a provided insights into experiments performed prior to computational modelling, the virtual screening on SMARCA4 and simulation of CYP3A4 guided subsequent experiments.

Chapter 2 of this thesis highlighted the *in silico* characterization of resorcine[4]arene cavitands. It was shown that modeled cavitand dynamics highly correlated with experiments and provided pivotal insights explaining previous inconsistencies in experimentally observed cavitand dynamics[1,2]. Emulation of the FRET experiment by a simple Markov state model generated from molecular dynamics distance and κ^2 values resulted in highly correlated donor fluorescence decay and fluorescence anisotropy decay curves. The results and correlation of simulation and experiment provides motivation for future application of molecular dynamics simulation for the design and dynamics analysis of small- to medium-sized organic compounds. Simple model systems like the cavitands, with their few degrees of freedom (compared to proteins), could also be employed to analyze dye-dye effects such as quenching, which were not included in the Markov state-based emulation presented herein. Given the high sensitivity of donor decay rates close to the Förster radius of the BODIPY dye, it appears crucial to generate the full protein-dye construct when modeling such systems. However, modeling a mixed system requires compatible small molecule and protein force fields; as such, development of missing parameters for residues linked to FRET dyes is critical.

Chapter 3 presented the characterization of cytochrome P450 3A4 (CYP3A4) in complex with carbamazepine (CBZ). The molecular dynamics simulations and subsequent clustering to obtain a putative carbamazepine binding mode guided mutational studies resulting in altered epoxidation kinetics for three mutants (A370V, A370L, S119A), one of which (A370V) displayed a significantly higher turnover rate at high CBZ concentrations. Another mutant (I369F) gave an increased turnover rate at low concentrations which are more consistent with generally prescribed doses of CBZ. The I369F CYP3A4 mutant was deemed most interesting for follow-up investigation by structural methods such as X-ray crystallography or NMR. In order to elucidate the cooperative behavior which was observed upon mutation, molecular dynamics simulations with two and more CBZ copies in the active site are indicative at best and should be subject to more rigorous simulation and evaluation by quantum mechanics-based analysis. In one of the simulations involving two CBZ molecules, a single CBZ left the active site through a channel previously described by Williams et al. 2004[3]. All suggested mutants showed increased ratios of ferric high-spin species, with S119A displaying the largest increase, emphasizing its critical role in substrate metabolism[4–9].

While chapter 3 presented the optimization of a protein towards a ligand, chapter 4 presents a more classical case of optimizing a library of ligands into a small set containing an active molecule against the SMARCA4 bromodomain. The virtual screening presented in chapter 4 employed an ALTA scheme incorporating tethered docking with rDock[10]. By applying this simple yet established screening technique we were able to reduce a multimillion database to a feasible set which included a 20 µM hit compound featuring a trans-glycoluril- N-[(2-hydroxyphenyl)methyleneamino]acetamide scaffold not previously reported as active on the SMARCA4 bromodomain. Next steps will include generating derivatives of this scaffold in order to assess the correctness of its putative binding mode, as well as optimizing this hit into a lead structure. X-ray crystallography of the protein complex with this scaffold is needed to confirm the binding mode. There is little evidence that bromodomains are specific to certain histone acetylation patterns. A specificity could be explained by combinatorial readout of multiple epigenetic marks[11]. Thus, considerations of the lack of specificity for acetylation patterns must be extended to future chemical probes and inhibitors, leading to high probability of cross-reactivity for any identified compound. Therefor it is highly recommended to perform a selectivity panel study to ensure that the herein presented compound is particularly active on SMARCA4.

The molecular dynamics simulations performed with iriomoteolide 3a and its target actin which were presented in chapter 5, suggest that the most probable binding site of iriomoteolide 3a is the barbed end. A multitude of chemically similar natural products which interfere with natural actin polymerization, have been structurally determined to bind at the barbed end, providing high confidence for the observed interaction site[12–15]. Binding to the barbed end of the actin monomer could explain the reduced actin polymerization by a monomer-sequestering mechanism. However, as it was observed that the two most sampled modes of binding do not interchange, insufficient sampling may limit the deduction of a binding mode. The different orientations of the macro-cycle indicate no specific interaction with the actin monomer aside hydrophobic interaction of the iriomoteolide 3a tail and the hydrophobic patch in the barbed end of actin. As this conflicts with the mode of interaction of other natural products, X-ray crystallography experiments are recommended to elucidate the nature of the complex.

Aggregation of organic molecules has been studied in the context of protein aggregation and micelle formation with and without proteins[16–18]. The observed aggregation of iriomoteolide 3a molecules in a high-concentration solution is therefore unsurprising. Aggregation of small molecules has been a longstanding problem in experimental studies and an understanding of the propensity of iriomoteolide 3a to aggregate is required in order to reconcile simulation with experiment[19].

Simulations with monomeric actin and bound reidispongiolide A showed that the macrocycle remained stable over five 100 ns simulations while in a single simulation the tail section dissociated before being reinserted into the barbed end. This observation (albeit sampled only once in five simulations) would support a model of actin fiber severing by insertion of its tail section into the inter-monomer region of two actin monomers after association of the macrocycle with the barbed end periphery[20]. However, as the event of redA tail dissociation happened only a single time, more sampling time would be required. Additionally, as redA is acting on F-actin, experimentally observed structures with the monomer are only limited in scope when trying to elaborate the exact mechanism of altered actin polymerization dynamics via reidispongiolide A.

References:

1. Azov VA, Schlegel A, Diederich F. Geometrically Precisely Defined Multinanometer Expansion/Contraction Motions in a Resorcin[4]arene Cavitand Based Molecular Switch. *Angew Chem Int Ed*. 2005;44(29):4635-4638. doi:10.1002/anie.200500970.

2. Azov VA, Beeby A, Cacciarini M, et al. Resorcin[4]arene Cavitand-Based Molecular Switches. *Adv Funct Mater*. 2006;16(2):147-156. doi:10.1002/adfm.200500181.

3. Williams PA, Cosme J, Vinković DM, et al. Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science*. 2004;305(5684):683-686. doi:10.1126/science.1099736.

4. He YA, He YQ, Szklarz GD, Halpert JR. Identification of Three Key Residues in Substrate Recognition Site 5 of Human Cytochrome P450 3A4 by Cassette and Site-Directed Mutagenesis. *Biochemistry (Mosc)*. 1997;36(29):8831-8839. doi:10.1021/bi970182i.

5. Park H, Lee S, Suh J. Structural and Dynamical Basis of Broad Substrate Specificity, Catalytic Mechanism, and Inhibition of Cytochrome P450 3A4. *J Am Chem Soc*. 2005;127(39):13634-13642. doi:10.1021/ja053809q.

6. Ekroos M, Sjögren T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc Natl Acad Sci*. 2006;103(37):13682-13687. doi:10.1073/pnas.0603236103.

7. Sevrioukova IF, Poulos TL. Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc Natl Acad Sci*. 2010;107(43):18422-18427. doi:10.1073/pnas.1010693107.

8. Sevrioukova IF, Poulos TL. Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. *Dalton Trans*. 2013;42(9):3116-3126. doi:10.1039/C2DT31833D.

9. Sevrioukova IF, Poulos TL. Dissecting Cytochrome P450 3A4–Ligand Interactions Using Ritonavir Analogues. *Biochemistry (Mosc)*. 2013;52(26):4474-4481. doi:10.1021/bi4005396.

10. Kolb P, Kipouros CB, Huang D, Caflisch A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins*. 2008;73(1):11-18. doi:10.1002/prot.22028.

11. Marchand J-R, Caflisch A. Binding Mode of Acetylated histones to Bromodomains: Variations on a Common Motif. *ChemMedChem*. 2015;10 ahead of print:1-8. doi:10.1002/cmdc.201500141.

12. Klenchin VA, King R, Tanaka J, Marriott G, Rayment I. Structural Basis of Swinholide A Binding to Actin. *Chem Biol*. 2005;12(3):287-291. doi:10.1016/j.chembiol.2005.02.011.

13. Rizvi SA, Tereshko V, Kossiakoff AA, Kozmin SA. Structure of Bistramide A−Actin Complex at a 1.35 Å Resolution. *J Am Chem Soc*. 2006;128(12):3882-3883. doi:10.1021/ja058319c.

14. Hirata K, Muraoka S, Suenaga K, et al. Structure Basis for Antitumor Effect of Aplyronine A. *J Mol Biol*. 2006;356(4):945-954. doi:10.1016/j.jmb.2005.12.031.

15. Klenchin VA, Allingham JS, King R, Tanaka J, Marriott G, Rayment I. Trisoxazole macrolide toxins mimic the binding of actin-capping proteins to actin. *Nat Struct Mol Biol*. 2003;10(12):1058-1063. doi:10.1038/nsb1006.

16. Marrink SJ, Tieleman DP, Mark AE. Molecular Dynamics Simulation of the Kinetics of Spontaneous Micelle Formation. *J Phys Chem B*. 2000;104(51):12165-12173. doi:10.1021/jp001898h.

17. Paci E, Gsponer J, Salvatella X, Vendruscolo M. Molecular Dynamics Studies of the Process of Amyloid Aggregation of Peptide Fragments of Transthyretin. *J Mol Biol*. 2004;340(3):555-569. doi:10.1016/j.jmb.2004.05.009.

18. Böckmann RA, Caflisch A. Spontaneous formation of detergent micelles around the outer membrane protein OmpX. *Biophys J*. 2005;88(5):3191-3204. doi:10.1529/biophysj.105.060426.

19. Rishton GM. Aggregator compounds confound amyloid fibrillization assay. *Nat Chem Biol*. 2008;4(3):159-160. doi:10.1038/nchembio0308-159.

20. Allingham JS, Zampella A, D'Auria MV, Rayment I. Structures of microfilament destabilizing toxins bound to actin provide insight into toxin design and activity. *Proc Natl Acad Sci U S A*. 2005;102(41):14527-14532. doi:10.1073/pnas.0502089102.

List of publications

Experimental and Computational Study of BODIPY Dye-Labeled Cavitand Dynamics

Igor Pochorovski, Tim Knehans, Daniel Nettels, Astrid M. Müller, W. Bernd Schweizer, Amedeo Caflisch, Benjamin Schuler, and François Diederich

J. Am. Chem. Soc. 2014, 136, 2441−2449

Concurrent Cooperativity and Substrate Inhibition in the Epoxidation of carbamazepine by Cytochrome P450 3A4 Active Site Mutants Inspired by Molecular Dynamics Simulations

Christian S. Müller, Tim Knehans, Dmitri R. Davydov, Patricia L. Bounds, Ursula von Mandach, James R. Halpert, Amedeo Caflisch, and Willem H. Koppenol

Biochemistry 2015, 54, 711−721

***In silico* identification of a trans-glycoluril scaffold interfering with the SMARCA4 bromodomain binding to acetylated lysine**

Tim Knehans, Amedeo Caflisch

to be submitted

Acknowledgements

Prof. Dr. Amedeo Caflisch was kind enough to take me in as a PhD student in 2011 in his highly diverse research group. I will never forget his support as my mentor and the lessons in science which he has taught me over the years.

I'd like to note my gratitude for the scientific support of the professors Ben Schuler, Cristina Nevado and Willem Koppenol.

I'd like to thank my collaborators Dr. Andrea Unzue-Lopez, Dr. Christian Müller, and particularly Dr. Igor Pochorovski for whom I have the utmost respect and will always remember the late nights we were working together on our shared interests in science or otherwise. I will always cherish the friendship which grew out of our collaborative efforts.

I want to thank all members of the Caflisch group who have helped me over the years in both moral and scientific terms. Special thanks go to Jean-Remi Marchand and Dr. Emilie Frugier who became close friends as well as Dr. Daniel Nettels of the Schuler group. Dr. Andreas Vitalis and the discussions we engaged in on a regular basis have been pivotal in accomplishing the here presented thesis. Thanks go also to the various people of the biochemistry institute like Yvonne Neldner, Dr. Cristina Batiste-Paulino who became close friends although very late in my PhD.

Finally, I would like to thank my parents Sabine and Jens Knehans as well as my sister Jana since without their relentless support this doctorate would have been much harder.

Curriculum Vitae

Tim Knehans

born 22nd August 1985 in Brandenburg an der Havel, Germany

from 39435 Wolmirsleben, Sachsen-Anhalt, Deutschland

Education

2011-2015	Graduate Studies: Computational Structural Biology / Biochemistry University of Zurich – Department of Biochemistry, 8057 Zurich, Switzerland
2005-2010	Undergraduate Studies: Master's degree (Diplom) Bioinformatics Johann Wolfgang Goethe Universität, 60438 Frankfurt am Main, Germany Master's Thesis: *Structure-based design of a flaviviral protease inhibitor* at DUKE-NUS Singapore, 169857 Singapore
1996-2005	High school Studies Gymnasium Egeln, 39435 Egeln, Germany